



The Indo-European Language Family

A Phylogenetic Perspective

Edited by Thomas Olander

The Indo-European Language Family

Modern languages like English, Spanish, Russian and Hindi as well as ancient languages like Greek, Latin and Sanskrit all belong to the Indo-European language family, which means that they all descend from a common ancestor. But how, more precisely, are the Indo-European languages related to each other? This book brings together pioneering research from a team of international scholars to address this fundamental question. It provides an introduction to linguistic subgrouping and offers comprehensive, systematic and up-to-date analyses of the ten main branches of the Indo-European language family: Anatolian, Tocharian, Italic, Celtic, Germanic, Greek, Armenian, Albanian, Indo-Iranian and Balto-Slavic. By highlighting that these branches are saliently different from each other, yet at the same time display striking similarities, the book investigates the early diversification of the Indo-European language family, spoken today by half the world's population. This title is also available as open access on Cambridge Core.

THOMAS OLANDER is Associate Professor in the Department of Nordic Studies and Linguistics at the University of Copenhagen. He is the author of *Balto-Slavic Accentual Mobility* (2009) and *Proto-Slavic Inflectional Morphology* (2015) and has edited several books on Indo-European linguistics and the linguistics–archaeology interface.

The Indo-European Language Family

A Phylogenetic Perspective

Edited by

Thomas Olander

University of Copenhagen



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108499798

DOI: [10.1017/9781108758666](https://doi.org/10.1017/9781108758666)

© Cambridge University Press 2022

This work is in copyright. It is subject to statutory exceptions and to the provisions of relevant licensing agreements; with the exception of the Creative Commons version the link for which is provided below, no reproduction of any part of this work may take place without the written permission of Cambridge University Press.

An online version of this work is published at doi.org/10.1017/9781108758666 under a Creative Commons Open Access license CC-BY-NC-ND 4.0 which permits re-use, distribution and reproduction in any medium for non-commercial purposes providing appropriate credit to the original work is given. You may not distribute derivative works without permission. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0>

All versions of this work may contain content reproduced under license from third parties. Permission to reproduce this third-party content must be obtained from these third-parties directly. When citing this work, please include a reference to the DOI [10.1017/9781108758666](https://doi.org/10.1017/9781108758666)

First published 2022

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Names: Olander, Thomas, 1974– author.

Title: The Indo-European language family : a phylogenetic perspective / edited by Thomas Olander.

Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2023. | This book has grown out of a workshop held in Copenhagen in February 2017, The Indo-European Family Tree. | Includes bibliographical references and index.

Identifiers: LCCN 2022032515 | ISBN 9781108499798 (hardback) | ISBN

9781108731522 (paperback) | ISBN 9781108758666 (ebook)

Subjects: LCSH: Indo-European languages. | BISAC: FOREIGN LANGUAGE STUDY / Ancient Languages (see also Latin) | LCGFT: Essays.

Classification: LCC P561 .O43 2023 | DDC 410–dc23/eng/20220727

LC record available at <https://lcn.loc.gov/2022032515>

ISBN 978-1-108-49979-8 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>List of Figures</i>	<i>page</i> vii
<i>List of Tables</i>	viii
<i>List of Contributors</i>	ix
<i>Acknowledgements</i>	x
<i>List of Abbreviations and Symbols</i>	xi
1 Introduction	1
THOMAS OLANDER	
2 Methodology in Linguistic Subgrouping	18
JAMES CLACKSON	
3 Computational Approaches to Linguistic Chronology and Subgrouping	33
DARIUSZ PIWOWARCZYK	
4 What We Can (and Can't) Learn from Computational Cladistics	52
DON RINGE	
5 Anatolian	63
ALWIN KLOEKHORST	
6 Tocharian	83
MICHAËL PEYROT	
7 Italo-Celtic	102
MICHAEL WEISS	
8 Italic	114
MICHAEL WEISS	
9 Celtic	135
ANDERS RICHARDT JØRGENSEN	

10	Germanic	152
	BJARNE SIMMELKJÆR SANDGAARD HANSEN & GUUS JAN KROONEN	
11	Greek	173
	LUCIEN VAN BEEK	
12	Armenian	202
	BIRGIT ANETTE OLSEN & RASMUS THORSØ	
13	Albanian	223
	ADAM HYLLESTED & BRIAN D. JOSEPH	
14	Indo-Iranian	246
	MARTIN JOACHIM KÜMMEL	
15	Balto-Slavic	269
	TIJMEN PRONK	
	<i>Index</i>	293

Figures

1.1	The “neo-traditional” model	<i>page</i> 5
1.2	Binary-branching model (Ringe group)	6
1.3	Binary-branching model (Gray group; Chang et al. 2015)	6
2.1	Family tree of the ABC language family	19
2.2	Schleicher’s tree diagram (Schleicher 1861: 7)	20
4.1	Outgroup analysis of PGmc. * <i>tūna</i> - and PCelt. * <i>dūn</i> . . .	53
5.1	The Luwic sub-branch of Anatolian	70
5.2	Tree model of the Anatolian languages	76
6.1	The position of Tocharian	89
7.1	Tentative tree showing the position of Italo-Celtic	109
8.1	The Italic languages	115
9.1	The Celtic languages	143
10.1	A tree model illustrating a binary split of Proto-Germanic into North-West vs. East Germanic	160
10.2	A unified tree-wave model of the Germanic dialect continuum	160
11.1	The Greek dialects	190
12.1	The position of Armenian	209
13.1	The position of Albanian	241
14.1	The Indo-Iranian languages	255

Tables

4.1	Number of rooted and unrooted binary branching trees	<i>page</i> 54
4.2	Best possible parsimony scores for West Germanic	57
10.1	Adjectival definiteness	155
14.1	Main phonological differences between Iranian and Indic	252
14.2	Morphological differences between Iranian and Indic	253
14.3	Morphological archaisms in Old Avestan	254
14.4	Morphological innovations in Old Avestan	254
14.5	Phonological changes in Iranian, Nuristanic and Indic	255
14.6	Examples of lexical differences between Iranian, Nuristanic and Indic	256

Contributors

LUCIEN VAN BEEK, Leiden University

JAMES CLACKSON, University of Cambridge

BJARNE SIMMELKJÆR SANDGAARD HANSEN, University of Copenhagen

ADAM HYLLESTED, University of Copenhagen

BRIAN D. JOSEPH, The Ohio State University

ANDERS RICHARDT JØRGENSEN, University of Copenhagen & Uppsala University

ALWIN KLOEKHORST, Leiden University

GUUS JAN KROONEN, Leiden University & University of Copenhagen

MARTIN JOACHIM KÜMMEL, Friedrich Schiller University Jena

THOMAS OLANDER, University of Copenhagen

BIRGIT ANETTE OLSEN, University of Copenhagen

MICHAËL PEYROT, Leiden University

DARIUSZ PIWOWARCZYK, Jagiellonian University in Kraków

TIJMEN PRONK, Leiden University

DON RINGE, University of Pennsylvania

RASMUS THORSØ, Leiden University

MICHAEL WEISS, Cornell University

Acknowledgements

This book is, of course, primarily the result of the hard work of the authors of the fifteen chapters. My main contribution as the editor of the book has been to remind the authors of upcoming and past deadlines and to ensure a certain degree of consistency across the chapters. I hope my efforts have resulted in a more coherent overall result.

I am very grateful to the authors for their willingness to contribute to the book and for their patience throughout the process. I also wish to express my gratitude to the editors at and around CUP – Andrew Winnard, Helen Barton, Isabel Collins, Subathra Manogaran and Kilmeny MacBride – for being more patient than one could possibly expect and for their willingness to discuss all kinds of editorial problems, small and large. Furthermore, I am thankful to Pete Westbrook for his meticulous language check of most of the chapters.

My work on this book was made possible by grants from the Carlsberg Foundation for the project *The homeland: In the footprints of the early Indo-Europeans* (2015–2018), from the Independent Research Fund Denmark for the project *Connecting the dots: Reconfiguring the Indo-European family tree* (2019–2023), and from Riksbankens jubileumsfond for the project *Languages and myths of prehistory: Unravelling the speech and beliefs of the unwritten past (LAMP)*. The Department of Nordic Studies and Linguistics, University of Copenhagen, has also facilitated my work on this book by providing a nice and pleasant work environment, which I truly appreciate. The most valuable part of my daily work environment is my colleagues at the *Roots of Europe* centre, who make every day at work enjoyable, a contribution that cannot be overestimated.

Finally, I am indebted to Jenny Helena Larsson, leader of the *LAMP* project, for generously providing the funding that has made it possible to publish this book open access.



Abbreviations and Symbols

abl.	ablative
acc.	accusative
act.	active
Aeol.	Aeolic
Alb.	Albanian
Anat.	Anatolian
Arc.	Arcadian
Arm.	Armenian
Att.	Attic
Av.	Avestan
BL	Bartholomae's Law
Boeot.	Boeotian
Br.	Brāhmaṇa
Brit.	Brittonic
Bulg.	Bulgarian
c.	common (gender)
Car.	Carian
Celtib.	Celtiberian
CS	Church Slavonic
CIr.	Common Indo-Iranian
CIrn.	Common Iranian
Cisalp.	Cisalpine
class.	classical
CLuw.	Cuneiform Luwian
coll.	collective
cpd.	compound
CPIE	Core Proto-Indo-European
Cr.	Croatian
CSl.	Common Slavic
CToch.	Common Tocharian
Cypr.	Cypriot
Cz.	Czech

dat.	dative
decl.	declension
dial.	dialect(al)
Dor.	Doric
du.	dual
EIA	Eastern Indo-Aryan
encl.	enclitic
Eng.	English
Ep.	Epic
ER	Early Runic
Eub.	Euboean
f(em).	feminine
Fal.	Faliscan
Far.	Faroese
Finn.	Finnish
Fr.	French
gen.	genitive
Gaul.	Gaulish
Geo.	Georgian
Gmc.	Germanic
Goth.	Gothic
Gr.	Greek
HG	High German
Hitt.	Hittite
HLuw.	Hieroglyphic Luwian
Hom.	Homer
Hsch.	Hesychius
Hung.	Hungarian
IA	(Eastern) Indo-Aryan
IE	Indo-European
Irn.	Indo-Iranian
Illyr.	Illyrian
impv.	imperative
Ind.	Indic
indic.	indicative
inf.	infinitive
instr.	instrumental
Ion.	Ionic
Ir.	Irish
Irn.	Iranian
Khot.	Khotanese
Lat.	Latin

Latv.	Latvian
len.	lenited
Lesb.	Lesbian
Lith.	Lithuanian
LLat.	Late Latin
loc.	locative
Luw.	Luwian
Lyc.	Lycian
MA	media(e) aspirata(e)
m(asc).	masculine
MBret.	Middle Breton
MCorn.	Middle Cornish
MDu.	Middle Dutch
Mess.	Messapic
MIA	Middle Indo-Aryan
mid.	middle
Mil.	Milyan
MIr.	Middle Irish
MIrn.	Middle Iranian
MLG	Middle Low German
ModBret.	Modern Breton
ModG	Modern German
ModGr.	Modern Greek
Mordv.	Mordvinian
MPers.	Middle Persian
MPhr.	Middle Phrygian
MW	Middle Welsh
Myc.	Mycenaean
neg.	negation
Neth.	Netherlandic (Dutch)
n(eut).	neuter
NFris.	North Frisian
NGeg	North Geg
NHG	New High German
NIA	New Indo-Aryan
NIE	Nuclear Indo-European
NIrn.	New Iranian
nom.	nominative
NPers.	New Persian
NPhr.	New Phrygian
Nur.	Nuristani(c)
Nw.	Norwegian

NWGr.	North-West Greek
OAlb.	Old Albanian
OAv.	Old Avestan
obj.	object
obl.	oblique
OCorn.	Old Cornish
OCS	Old Church Slavonic
OCz.	Old Czech
ODu.	Old Dutch
OE	Old English
OF	Old Frisian
OFr.	Old French
OHG	Old High German
OIA	Old Indo-Aryan
OIr.	Old Irish
OIrn.	Old Iranian
OLat.	Old Latin
OLith.	Old Lithuanian
ON	Old Norse
OPers.	Old Persian
OPhr.	Old Phrygian
OPr.	Old Prussian
ORu.	Old Russian
OS	Old Saxon
Osc.	Oscan
OSw.	Old Swedish
OUmb.	Old Umbrian
Pael.	Paelignian
PAlb.	Proto-Albanian
PAnat.	Proto-Anatolian
PBalt.	Proto-Baltic
PBret.	Proto-Breton
PBrit.	Proto-Brittonic
PCelt.	Proto-Celtic
perl.	perlative
pf.	perfect
PGmc.	Proto-Germanic
Phryg.	Phrygian
PIA	Proto-Indo-Aryan
PIE	Proto-Indo-European
PIIrn.	Proto-Indo-Iranian
PIrn.	Proto-Iranian

Pis.	Pisidian
Plt.	Proto-Italic
PNIE	Proto-Nuclear Indo-European
Port.	Portuguese
pl.	plural
pres.	present
pret.	preterite
ps.	person
PSl.	Proto-Slavic
ptc.	participle
PToch.	Proto-Tocharian
PWGmc.	Proto-West-Germanic
Ru.	Russian
RuCS	Russian Church Slavonic
Samn.	Samnite
SCr.	Serbo-Croatian
SE	secondary ending
Serb.	Serbian
sg.	singular
sglt.	singulative
SGr.	South Greek
Sid.	Sidetic
Skt.	Sanskrit
Slk.	Slovak
Sln.	Slovene
Sogd.	Sogdian
Sp.	Spanish
SPic.	South Picene
SWBrit.	South-West Brittonic
SWNw.	South-West Norwegian
Thess.	Thessalian
Thrac.	Thracian
Toch.	Tocharian
Toch.A	Tocharian A
Toch.B	Tocharian B
Umb.	Umbrian
Ved.	Vedic Sanskrit
Ven.	Venetic
VOL	Very Old Latin
W	Welsh
WFris.	West Frisian
WGr.	West Greek

WIA	Western Indo-Aryan = Mittanni (Indo-)Aryan
YAv.	Young Avestan
C	consonant
H	laryngeal
R	sonorant consonant
V	vowel
*	reconstruction
< >	phonetically regular change
← →	morphological reshaping or replacement
⇐ ⇒	borrowing

1 Introduction

Thomas Olander

1.1 Background

The study of the genealogical relationship between the Indo-European languages has been the object of research ever since August Schleicher's famous Stammbaum representation of the then-known subgroups, or branches (1861: 7; see also 1853: 787). Throughout most of the twentieth century, this topic played a less prominent role in Indo-European studies, but the last few decades have witnessed a surge of interest in the internal structure of the Indo-European language family as well as other language families.

From a methodological point of view, the renewed interest in linguistic phylogenetics, or “phylolinguistics”, came mainly from two sides, rather different in their choice of methods and data, yet both based on computational approaches. A group of researchers led by Don Ringe applied algorithms based on weighted maximum compatibility to a data set consisting of phonological and morphological characters and a list of basic vocabulary items from a selection of twenty-four Indo-European languages representing the individual subgroups (Ringe, Warnow & Taylor 2002; Nakhleh, Ringe & Warnow 2005). Another group, headed by Russell D. Gray, applied Bayesian methods to data sets exclusively consisting of lists of basic vocabulary (for the Indo-European language family, see e.g. Gray & Atkinson 2003; Bouckaert et al. 2012); the same methods and data were used in Chang et al. 2015.

Within Indo-European studies, the increasing interest in linguistic phylogenetics has mainly taken its point of departure in traditional methodology, where subgroups are identified on the basis of significant shared innovations across related languages. It seems likely that specialists have become more interested in the branching structure of the family tree as a result, at least partly, of the growing acceptance of the Anatolian subgroup as a sister to all the remaining

This chapter was written in connection with the research projects *Connecting the dots: Reconfiguring the Indo-European family tree* (2019–23), financed by the Independent Research Fund Denmark, and *LAMP: Languages and myths of prehistory* (2020–5), financed by Riksbankens Jubileumsfond. I am grateful to Simon Poulsen for reading and commenting on a draft version of the chapter.

Indo-European languages (see e.g. [Kloekhorst 2008: 7–11](#); [Kloekhorst & Pronk 2019](#); [Oettinger 2014](#); but cf. the more sceptical stance by [Melchert in press](#)), which highlights the importance of the structure of the family tree for the purposes of reconstruction.

This book has grown out of a workshop held in Copenhagen in February 2017, “The Indo-European Family Tree”, where invited speakers discussed methodological issues and the phylogenetic relations of each of the main Indo-European subgroups. Some of the chapters of this book have been authored by participants in that workshop, while others have been written by authors invited to contribute to the book project.

The Copenhagen workshop was organised within the framework of the research project *The homeland: In the footprints of the early Indo-Europeans* at the University of Copenhagen (2015–18, financed by the Carlsberg Foundation). The *Homeland* project was concerned with the location in time and space of the speakers of Proto-Indo-European and the early spread of the Indo-European language family throughout Europe and western Asia. Since the nodes of a linguistic family tree to a certain extent historically represent the geographical separation of the speakers, it is essential, when attempting to correlate prehistoric languages with material culture, to have a good understanding of the order of separation of the daughter languages from their common ancestor. Thus, the so-called Indo-European homeland problem and the problem of the structure of the Indo-European family tree are closely intertwined. Indeed, studies of linguistic phylogenetics are very often also concerned with the geography and time depth of the nodes in the tree, even if the methodologies involved are very different (for Indo-European see e.g. [Nakhleh, Ringe & Warnow 2005](#) and [Bouckaert et al. 2012](#)).

In its design and structure this book is rooted in the traditional methodology of linguistic subgrouping. This is not only because of the background to how the book was conceived. Over the last couple of decades, computer-assisted approaches have, in my view, received more attention than can be justified by the results they have produced. In some circles, especially within non-linguistic disciplines and among a broader audience (as exemplified by the media coverage of [Bouckaert et al. 2012](#)), computer-assisted approaches seem to be more highly regarded than traditional studies.

The impact of publications based on computer-assisted approaches has been very limited within the field of Indo-European studies itself, although the results achieved by the Ringe group have been somewhat successful (see [Clackson 2015: 5](#)). Interestingly, what we see is not a large-scale rejection of the findings of computer-assisted approaches by traditional Indo-European linguists. The findings are, in most cases, simply ignored,

probably due to a combination of factors, including the fact that computational phylogenetic studies are difficult to evaluate for non-computational linguists. This is because the methods employed are very different from traditional methods in a number of ways. Firstly, the main focus of computational studies is often on the methodology and the results, rather than on the actual data, which are often full of errors. Secondly, computational studies are often written in a very technical language. And thirdly, the results are not thought to be of any actual value anyway as they are often based on material that is not considered to be particularly significant, while the most relevant material is ignored.

Thus, in some sense, this book may be seen as a traditionalist reaction to modern computer-assisted approaches to linguistic Indo-European phylogenetics. This does not mean, however, that the contributors to the book in any way have ignored the fact that such approaches may be of great benefit to linguistic phylogenetics in general or to Indo-European studies in particular; see the chapters by Clackson ([Chapter 2](#)), Piwowarczyk ([Chapter 3](#)) and Ringe ([Chapter 4](#)). What should be evident from the book is that traditional approaches still have a lot to offer, even though they require a high degree of specialisation, including a deep understanding of the comparative method and linguistic reconstruction as well as a profound knowledge of the relevant data, constituted by the phonology, morphology, syntax and lexicon of a large number of languages and their historical development from Proto-Indo-European to their attestation. As is often emphasised, computational methods cannot and should not replace traditional historical linguistics but may prove to be a useful supplement (Ringe, Warnow & Taylor 2002: 65–6; compare also the very enthusiastic remarks on Bayesian linguistic phylogenetics by Greenhill, Heggarty & Gray 2021: 246 with the critical position by Ringe in [Chapter 4](#) of this book). This book is thus, in some way, an attempt at reinvigorating the traditional methodology, which, outside Indo-European studies, seems to be losing ground to computationally based analyses.

In traditional Indo-European linguistics, there are surprisingly few comprehensive studies of the phylogeny of the language family. Two works that were influential in their time are Antoine Meillet's *Les dialectes indo-européens* (1908/1922) and Walter Porzig's *Die Gliederung des indogermanischen Sprachgebiets* (1954). Both works are now old and outdated in a number of respects, and perhaps more importantly, their primary aim is to analyse the relationship between the ancient Indo-European languages with respect to their geographical location rather than to uncover the phylogenetic structure of the Indo-European family tree.

Somewhat newer, but still more than half a century old, is *Ancient Indo-European dialects*, edited by Henrik Birnbaum and Jaan Puhvel (1966). While

some parts of that book, especially those concerned with methodological problems, are still useful, and some of the chapters even have similar titles to those found in this book, it does not cover the individual subgroups systematically but only highlights some aspects. Like Meillet's and Porzig's books, it is also outdated in a number of respects.

Up-to-date from the point of view of Indo-European linguistics, the work by the Ringe team (e.g. Ringe, Warnow & Taylor 2002; Nakhleh, Ringe & Warnow 2005) is partly based on the traditional methodology in that it identifies significant shared innovations; in addition it incorporates shared basic vocabulary items. In contrast to traditional Indo-European linguistics, the Ringe team uses computational methods to produce the best family tree based on a weighted algorithm. Since the work by the Ringe team has been published in articles and book chapters, rather than in book-length treatments, it does not offer much in the way of extensive qualitative discussion of the evidence provided by the individual subgroups. One of the aims of this book is to facilitate this kind of discussion.

It may be worthwhile to ask why the structure of the Indo-European family tree attracts so much interest. For specialists it is essential to have an idea of the branching structure of the family tree in order to arrive at an adequate reconstruction of the Indo-European proto-language and its development into the attested Indo-European languages. All language-internal aspects of reconstructed Proto-Indo-European – phonology, inflectional and derivational morphology, syntax, lexicon – depend on the relationship between the individual subgroups. Any linguistic feature – say, the phoneme **b*, the augment or the word for 'wheel' – must be viewed in the light of the family tree (Olander 2018). If the feature cannot be reconstructed back to Proto-Indo-European itself, it may or may not have been present in the proto-language, but the phylogenetic information should be included in the evaluation of each feature, along with systemic and typological considerations and the evidence of internal reconstruction.

Other aspects of Indo-European studies are also intimately connected with the purely linguistic evidence. For instance, as already mentioned, the branching structure is very likely to be related to the geographical spread of early Indo-European speech varieties, and the existence of terminology for concepts like 'wheel' in the proto-language of a given linguistic subgroup is crucial for pinpointing the geographical and chronological location of that proto-language. Thus, in correlating the Indo-European proto-language and the prehistoric spread of Indo-European languages with the archaeological record – including the identification of the Indo-European homeland – the branching structure of the family tree plays a decisive role. As this question has appeal that goes well beyond specialist circles, the branching structure of

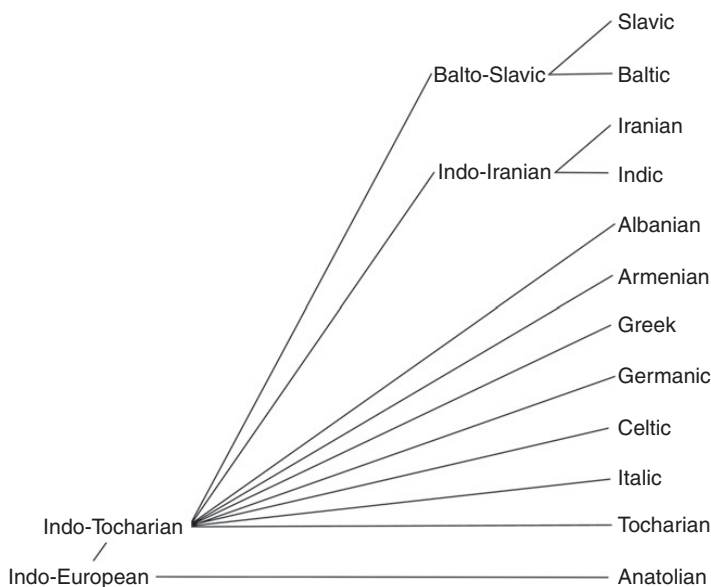


Figure 1.1 The “neo-traditional” model

the family tree is not only highly significant in the field of Indo-European studies but has a great impact on a broader audience as well.

The following illustrations show some of the models of the Indo-European language family that can be found in recent publications (the nodes are named according to the suggestion in [Olander 2019a](#)). First, though rarely made explicit, the tree underlying much work in Indo-European studies is the “neo-traditional model”, where the Anatolian subgroup separates first, whereas the relationship between the remaining subgroups is undetermined, de facto resulting in a non-hierarchical subtree for the non-Anatolian part of the family; see [Figure 1.1](#).

A radically different structure is assumed by the Ringe group. The tree is binary-branching, with the subgroups leaving gradually; see [Figure 1.2](#) (based on [Nakhleh, Ringe & Warnow 2005](#): 397, tree 5A). The position of Albanian in this tree is uncertain.

The Gray group also works with a binary-branching tree, but one that differs from the previous one except in the initial splits ([Bouckaert et al. 2012](#), with a revised tree in [Bouckaert et al. 2013](#)); see [Figure 1.3](#). The same tree is found in [Chang et al. 2015](#).

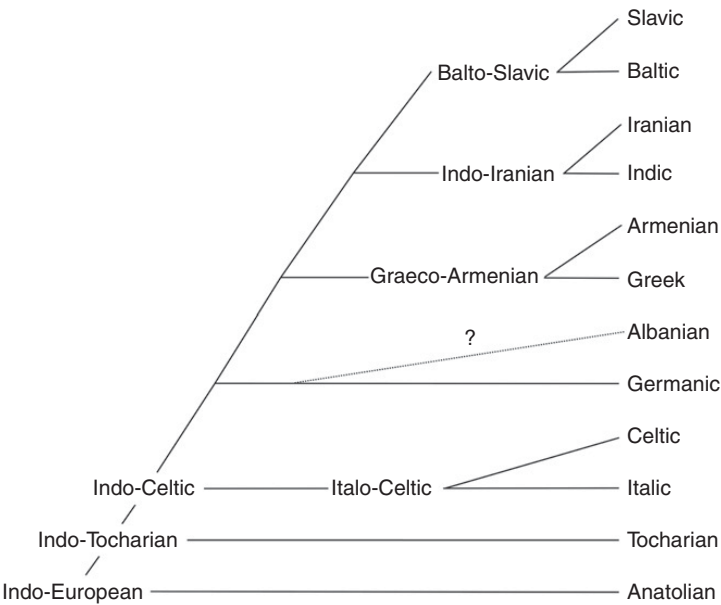


Figure 1.2 Binary-branching model (Ringe group)

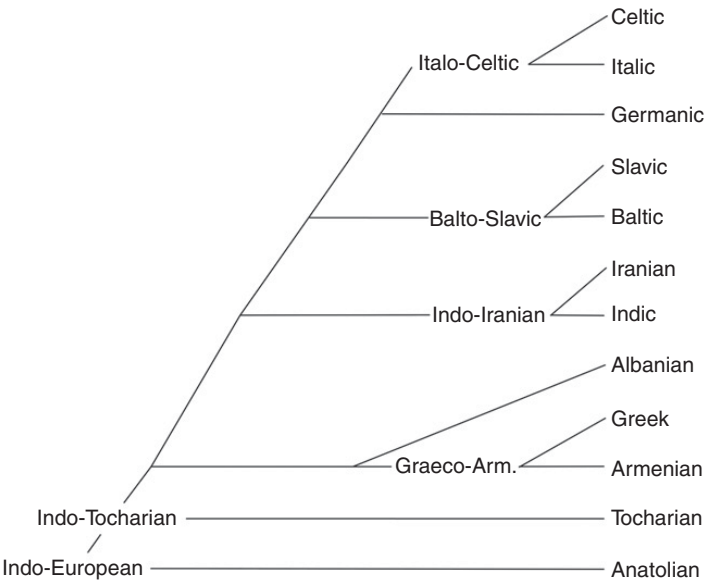


Figure 1.3 Binary-branching model (Gray group; Chang et al. 2015)

1.2 Terminology

If authors use the same terms for different phenomena, misunderstandings easily arise, especially across different disciplines. Therefore, I wish to explore in some detail a term that is a recurring topic for discussion in historical linguistics yet which still causes much confusion, namely *proto-language*, a central concept in phylogenetic linguistics and in discussions of linguistic homelands. Most linguists would agree that the term refers to the last common ancestor of a group of related languages (see the discussion in Olander 2019b: 10–12), but since “the last common ancestor” means different things to different authors, there is often little actual agreement on the content.

In works based on cognacy databases, including Bayesian studies, I have not seen an explicit definition of the concept of a proto-language. However, as long as all items in the basic vocabulary lists of two or more speech varieties are cognate, these varieties are still considered to be one language. Accordingly, I assume, a proto-language does not dissolve as long as no word in the list is replaced by another word in one of the varieties. This mechanism may lead to undesired results. To give an exaggerated example for illustrative purposes, we might hypothetically assume two related speech varieties where all the basic words are cognate, but where, apart from that, there is only a minimal lexical overlap between the two varieties. Moreover, the varieties have diverged significantly phonologically and morphologically; for instance, one variety has [o] and [ʔəvø] for ‘water’ and ‘hair’, with the cognates [ˈakkwa] and [kaˈpelli] in the other. The nominal and verbal inflectional systems are very simple in the former variety, while the latter has a nominal system with numerous cases and an elaborate verbal system with several tenses, aspects and moods. These varieties would still be considered one uniform entity in frameworks that only take basic vocabulary into consideration.

In traditional historical linguistics, by contrast, a proto-language usually refers to the stage of a language immediately before the first linguistic change – not only in the basic vocabulary – that does not affect all daughter languages (cf. Eichner 1988: 11–20; Olander 2015: 18–21 with references). By this definition a proto-language is a uniform entity with no dialects or other varieties. It is clear that this somewhat idealised definition, which refers to only one speech variety, does not correspond to a “real” language, which usually comprises a number of varieties. However, the definition is unambiguous and, crucially, a proto-language is the result of the application of the comparative method to a set of related languages, which makes very good sense from the point of view of historical linguistics.

Still from the point of view of traditional historical linguistics, it may also be useful to be able to refer to a group of related speech varieties that have already diverged from each other yet are still close enough to introduce identical or

near-identical innovations. While some authors may conceive this as a proto-language, I prefer to reserve that term for the above-mentioned concept and to use *common language* to refer to this latter concept (cf. Olander 2015: 18–21 for the general terminology, and 29–31 for its application to Slavic). Applying these definitions, then, Proto-Indo-European is the stage before the first linguistic change in any speech variety, whereas Common Indo-European refers to a group of already differentiated Indo-European varieties that are still linguistically close enough to carry out common innovations.

In terms of absolute chronology, the stage immediately before any linguistic change in the speech community (detectable by the comparative method) logically precedes, usually by a considerable amount of time, both the last stage where common innovations are still possible and the stage immediately before a lexical item is replaced on a basic vocabulary list. Thus apparently similar ways of defining a proto-language (“last common ancestor”) may, if understood differently, lead to widely diverging results. When taking into consideration how significant this terminological discrepancy may be, it is rather surprising that it is only very rarely addressed in the literature.

Since the *homeland* of a proto-language is, to most authors, the location in space and time where a given proto-language was spoken (cf. Eichner 1988: 20–1; Olander 2019b: 10–12), a precise understanding of what a proto-language refers to is central in discussions of linguistic homelands. If different definitions of a proto-language end up identifying language stages separated by several centuries or even millennia, it is not surprising that there is disagreement on when and where these stages were spoken. It is, in my view, quite possible that some of the disagreement about the time and place of the Indo-European homeland is directly caused by this terminological confusion. I should add that, in my opinion, a linguistic homeland should, for practical purposes, refer to the location in time and space where a common language, not a proto-language, was spoken (in the sense of the words just discussed).

Another term that may be useful to introduce in discussions of proto-languages is a *para-proto-language*, which refers to the related speech varieties spoken at the same time as a given proto-language. For instance, Proto-Indo-European as we reconstruct it using the comparative method is one variety among several varieties spoken at the same time; these varieties may be referred to as Para-Proto-Indo-European. While we do not know much about these para-languages, which have subsequently been displaced by other speech varieties, their earlier presence may be indicated, e.g., by phonological irregularities in words that are apparently inherited from Proto-Indo-European but which may actually have been borrowed from Para-Proto-Indo-European varieties.

If we accept that Anatolian and perhaps Tocharian were the two first subgroups to separate (see the [next section](#)), then there must have existed

intermediate proto-languages below the level of Proto-Indo-European but above the level of the proto-languages of the individual subgroups – for instance the proto-language of the non-Anatolian subgroups, that of the non-Anatolian and non-Tocharian subgroups, as well as that of Italic and Celtic and that of Greek and Armenian. The need to be able to designate these intermediate proto-languages has been highlighted in Olander 2019a (see also the careful considerations on the interpretation of a family tree, including the internal nodes, by Ringe, Warnow & Taylor 2002: 109; but cf. the provocative statement by Garrett 1999: 147 that “the intermediate nodes ... are nameless precisely because we do not need to refer to them”). I have applied the terminological principles laid out in Olander 2019a to the figures of the present chapter.

It is important to acknowledge that these intermediate proto-languages are not defined by being residual compared to the subgroups they do not include. On the contrary, they are posited precisely because the subgroups descending from them display shared innovations, unlike the remaining subgroups (cf. Ross 1997: 222). If no shared innovations can be shown for a suggested intermediate proto-language, that proto-language is not justified in the model.

1.3 Contents and Structure of the Book

The book contains fifteen chapters. The first four chapters outline the background to the book and address methodological issues. They also deal, from different perspectives, with the question of what the book is not, by discussing recent computational approaches to linguistic phylogenetics and why they are problematic.

In this introductory chapter, the background and motivation for the book are outlined, and some of the terminological issues pertaining to linguistic reconstruction and linguistic phylogenetics are addressed. It summarises the content of the remaining chapters and discusses some of the perspectives they raise.

Chapter 2, “Methodology in Linguistic Subgrouping” by James Clackson, shows how scholars have discussed the phylogeny of the Indo-European language family for the last 200 years, and it sets out the methodological choices that face current and future researchers. Since the late nineteenth century, it has been generally agreed that the best supporting evidence for a subgroup of A and B is the existence of non-trivial shared linguistic innovations made in both A and B but not in C. There is, however, still debate as to what counts as non-trivial, how to identify shared innovations that arose through language contact, how many innovations are required to construct a subgroup, and whether splits are necessarily binary. These debates are further explained and explored in the chapter.

Chapter 3, “Computational Approaches to Linguistic Chronology and Subgrouping” by Dariusz Piwowarczyk, presents an overview of computer-assisted approaches to linguistic subgrouping, highlighting advantages and drawbacks of the individual methods and evaluating the results achieved by applying these approaches. Since the exact same set of changes in the same order in two languages can be a sign of common development and, accordingly, of a subgroup, the chapter pays special attention to the potential of computational simulations of sound change. This approach is illustrated by material drawn from different subgroups thought to be closely related, starting from the most obvious ones (Indo-Iranian) to the ones that are less obvious (Balto-Slavic) and even controversial (Italo-Celtic, Graeco-Armenian).

Chapter 4, “What We Can (and Can’t) Learn from Computational Cladistics” by Don Ringe, investigates the advantages and limitations of computational approaches to linguistic phylogenetics. It discusses the intractable size of cladistic data sets, which can only be processed using computational methods, the relative unreliability of lexical data, and the ways in which phonological and inflectional data must be used together to construct and root a cladistic tree. It also considers how to handle language groups with only partly treelike diversification. Finally, the chapter critiques some recent high-profile cladistic analyses from several angles, exposing further pitfalls in the incautious use of cladistic tools. Its conclusions are only moderately positive, but are argued to be realistic.

The remaining eleven chapters each deal with one of the major Indo-European subgroups: Anatolian, Tocharian, Italic, Celtic, Germanic, Greek, Armenian, Albanian, Indo-Iranian and Balto-Slavic, plus the putative Italo-Celtic subgroup. Fragmentarily documented subgroups such as Phrygian and Messapic are not treated separately, but their positions in the family tree are discussed in relation to the major subgroups. The chapters have a similar structure. Each subgroup is presented together with its attestation, geographical distribution etc., the evidence for the subgroup, its internal subgrouping, its relationship to the other subgroups and a discussion of the position of the subgroup in the overall family tree of Indo-European. Since the subgroups are very different from each other on the various parameters, the chapters focus on different aspects of the phylogenetic description. For instance, as the Italic subgroup is more diversified by its earliest attestation than Armenian, the section dealing with the internal structure of Italic (Section 8.3) is much more comprehensive than the corresponding Armenian section (Section 12.3).

Chapter 5, by Alwin Kloekhorst, presents the Anatolian languages and some of their prominent linguistic features, discussing whether they represent archaisms or innovations, only the latter being indicative of an Anatolian subgroup. The chapter proceeds with an analysis of the internal subgrouping of the

Anatolian languages, arguing for a Hittite subgroup and a subgroup comprising Lydian, Palaic and the Luvic languages. After a review of the alleged “western” affinities of Anatolian, the chapter discusses one of the most prominent problems in Indo-European phylogenetics over the last several decades, namely the question of whether the Anatolian subgroup was the first one to separate from the remaining Indo-European languages. It concludes that Anatolian is indeed the outlier in the family, and that the gap between the split-off of Anatolian and the rest is substantial.

Chapter 6, by Michaël Peyrot, introduces the two closely related languages known as Tocharian A and Tocharian B. It addresses the most important shared innovations that characterise these languages and thus define the Tocharian subgroup. This is followed by an analysis of the genealogical relationship with the other subgroups, especially Anatolian. The chapter also assesses the position of Tocharian in the Indo-European family tree, where Tocharian is often considered to be the second subgroup to separate, and reviews the arguments for and against this hypothesis. It is concluded that the question is still open, to some extent because the overall structure of the Proto-Indo-European verbal system is uncertain, which makes it difficult to distinguish innovations from archaisms in the descendants, including Tocharian.

Chapter 7, by Michael Weiss, contains two main subsections. The first one discusses the reality of an Italo-Celtic subgroup within the Indo-European language family, concluding that there is enough evidence to assume a genuine but short-lived subgroup. The second subsection analyses the overall position of Italo-Celtic in the family tree.

Chapter 8, also by Michael Weiss, offers a presentation of the Italic subgroup, the reality of which has sometimes been called into question, although it seems to be supported by a substantial number of shared innovations. The chapter addresses the internal subgrouping of Italic, where Latin and Faliscan constitute one subgroup, and Oscan and Umbrian another, the position of Venetic being unclear. The relationship between Italic and the other subgroups (except Celtic; see above) is discussed.

Chapter 9, by Anders Richardt Jørgensen, first presents the Celtic languages and discusses the arguments, mostly of phonological nature, for a Celtic subgroup. The internal subgrouping of Celtic is contested: while the existence of a Goidelic and a Brittonic subgroup is uncontroversial, it is uncertain whether Brittonic forms a subgroup with Gaulish, with Goidelic or with neither of them. The chapter discusses the relationship of Celtic with Germanic and the other subgroups (except Italic; see above).

Chapter 10, by Bjarne Simmelkjær Sandgaard Hansen and Guus Jan Kroonen, introduces the Germanic languages, listing the most salient features characterising that subgroup. The chapter discusses the relationship between East, West and North Germanic, concluding that the latter two subgroups are

more closely related to each other than to the former. The subgroups that seem to be most closely associated with Germanic are Italic, Celtic and Balto-Slavic, although none of them appears to form an actual subgroup with Germanic in the family tree. Despite being innovative in many respects, Germanic also preserves certain archaic features that suggest it may have been one of the first subgroups to separate from the core group.

Chapter 11, by Lucien van Beek, presents the Greek subgroup, arguing for its reality based on several innovations found in all varieties of Greek. It addresses the complicated question of the internal subgrouping of Greek and the relationship of Greek to Macedonian, Phrygian and Armenian, concluding that Macedonian may possibly be classified as a Greek dialect and that Phrygian constitutes a subgroup with Greek. The relationship between Armenian and Greek is not as close as is often maintained (cf. **Chapter 12**). The position of Graeco-Phrygian in the family tree, and especially the relationship with Indo-Iranian, is also discussed.

Chapter 12, by Birgit Anette Olsen and Rasmus Thorsø, examines Armenian, listing the innovations that constitute the evidence for the reality of the Armenian subgroup. It then analyses the relationship of Armenian to other subgroups of Indo-European, first of all Greek, but also Phrygian and Albanian, arguing that Armenian constitutes a higher-order subgroup, “Balkanic”, together with these three subgroups. Within the Balkanic group, Greek and Phrygian are most closely related, and together with Armenian they constitute a larger subgroup. Armenian and Albanian, on the other hand, do not share any exclusive innovations within Balkanic.

Chapter 13, by Adam Hyllested and Brian D. Joseph, gives an overview of Albanian. After a brief discussion of the features that constitute Albanian as a separate subgroup and a presentation of the dialect divisions within Albanian, the chapter analyses the relationship of Albanian to the other subgroups. Special attention is given to the relationship between Albanian and Greek, which are regarded as forming a subgroup within a Balkanic group also consisting of Armenian and Phrygian.

Chapter 14, by Martin Joachim Kümmel, presents the Indo-Iranian subgroup, discussing the relationship between Indic and Iranian and assessing the difficult question of the position of the Nuristani languages. It analyses the position of Indo-Iranian within the Indo-European family tree, arguing that it may have separated relatively early and stayed in contact with several other subgroups.

Chapter 15, by Tijmen Pronk, covers the Baltic and Slavic languages. It analyses the much-debated relationship between the two groups, concluding that they do constitute a subgroup together. The chapter discusses the question of the internal structure of Balto-Slavic, especially the position of Old Prussian between East Baltic and Slavic, and it analyses the relationship of Balto-Slavic

to Germanic and Indo-Iranian, arguing that Balto-Slavic does not form a higher-order subgroup with these or other subgroups.

1.4 Results and Perspectives

As should be all too clear from the [preceding section](#), this book does not solve all problems related to the higher-order phylogeny of the Indo-European language family. On the contrary, in many respects it raises more questions than it answers. At the same time, it also highlights the necessity not only of examining in more detail individual potentially shared innovations across subgroups but also of zooming out and looking at the entire family, and the importance of methodological considerations. The latter question is the topic of the chapters by Clackson ([Chapter 2](#)), Piwowarczyk ([Chapter 3](#)) and Ringe ([Chapter 4](#)), who investigate different methodological aspects of linguistic phylogenetics.

With the exception of Balto-Slavic and, to a lesser extent, Italic, the reality of the main subgroups is hardly ever called into question; the rare exceptions have not found much support in the scholarly community (e.g. the doubts about Greek being a subgroup expressed by [Garrett 2006](#); cf. also the characterisation of Iranian as a Sprachbund by [Tremblay 2005](#): 687). In this book, the similarities between Baltic and Slavic are considered to be so striking that they are dealt with together in one chapter ([Chapter 15](#) by Pronk). Similarly, the Italic languages display enough common innovations that they are also regarded as a real subgroup of Indo-European ([Chapter 8](#) by Weiss). Considerably less certain is a group consisting of Italic and Celtic; in this book the relationship between these subgroups is considered sufficiently important to merit a chapter on Italo-Celtic ([Chapter 7](#) by Weiss), although Italic and Celtic are also discussed in separate chapters ([Chapter 8](#) by Weiss and [Chapter 9](#) by Jørgensen, respectively).

The view that Indic and Iranian, while clearly separate subgroups, do form a subgroup together is unchallenged, although the position of Nuristani within Indo-Iranian is still disputed, and there is no agreement on the position of Indo-Iranian in the overall family tree ([Chapter 14](#) by Kümmel).

When it comes to the higher-order grouping, however, the situation is much less straightforward. A recurring theme throughout the book is that most of the individual subgroups are very difficult to place in the overall family tree, except Anatolian, for which the idea of an early separation has gained much traction in recent decades and is further supported in this book by Kloekhorst ([Chapter 5](#)). The position of Tocharian, often regarded as the second subgroup to separate, cannot be established with any certainty since shared innovations of the remaining subgroups are difficult to determine, as argued by Peyrot ([Chapter 6](#)).

Weiss (Chapter 7) discusses the idea that Italo-Celtic may have split off relatively early from the tree, perhaps after Tocharian. Germanic, showing affinities above all with Balto-Slavic and Italic, is difficult to place in the overall tree, as argued by Hansen and Kroonen (Chapter 10). The mutual relationship between the “Balkanic” languages – Greek (Chapter 11), Armenian (Chapter 12), Albanian (Chapter 13) as well as scantily attested languages such as Phrygian and Messapic – is evaluated differently by the authors of this book. While Greek is thought to constitute a phylogenetic unit together with Phrygian in all three chapters, the hypothesis of a Graeco-Armenian subgroup is given a negative appraisal by van Beek (Chapter 11), while Olsen and Thorsø (Chapter 12) are positive. A third position is taken by Hyllested and Joseph (Chapter 13), who argue that Greek forms a subgroup with the notoriously difficult Albanian. Interestingly, the evidence for a subgroup consisting of Indo-Iranian and Balto-Slavic, occasionally discussed in the literature (Søborg 2020: 52; cf. Ringe, Warnow & Taylor 2002: 103–4), is considered to be insufficient by both Kümmel (Chapter 14) and Pronk (Chapter 15).

Even without decisive answers to many of the questions that were also being asked in Indo-European linguistic phylogenetics a decade and a half ago, these diverging conclusions – rather than indicating that the endeavour of modelling the Indo-European family tree is a failure – contribute to a more diverse picture of the dissolution of the Indo-European proto-language. When the evidence is not clear-cut, it is natural that assigning different weight to the various pieces of evidence leads to different conclusions. Interestingly, the different conclusions reached in the various chapters only rarely seem to hinge on discrepancies in the reconstruction of Proto-Indo-European and its development into the individual daughter languages, although one might have expected such discrepancies to play a significant role.

This book examines the Indo-European language family from the point of view of each of the ten main subgroups of Indo-European. While a systematic individual assessment of the subgroups is an indispensable first step towards a better understanding of the internal structure of the Indo-European family tree, it is also clear that there is a need for a systematic reassessment of the Indo-European family tree as a whole. It remains an open question as to whether this should be done purely by applying the traditional methodology, which seeks to identify and evaluate significant shared innovations, or if computational methods, which make it possible to work with larger data sets, can contribute anything of true value.

If we are able to obtain a relatively solid picture of the higher-order subgrouping of the Indo-European language family, the family tree may serve as a vital means of solving problems of Indo-European reconstruction. Any

reconstruction should be evaluated in the light of the family tree, and a reconstruction suggested by several subgroups is only justified for Proto-Indo-European itself if it is compatible with the outlier in the family (see [Ringe 1998](#); [Olander 2018](#)). Without an understanding of the structure of the Indo-European family tree, it is also difficult to trace the prehistoric spread of the Indo-European languages throughout Europe and western Asia.

1.5 Practical Remarks

I have strived to harmonise the notation of attested and reconstructed forms in the individual chapters without forcing my own views on the authors. For instance, the purely conventional use of $*\hat{k} \hat{g} \hat{g}^h$ and $*\hat{j} \hat{u}$ has been introduced for $*\hat{k} \hat{g} \hat{g}^h$ and $*y w$ in Proto-Indo-European reconstructions. However, when the notational differences are the result of different conceptions of the reconstructed forms, I have retained the authors' preferences. Thus I have not harmonised e.g. the absence or presence of laryngeal colouring (e.g. pf.1sg. $*\hat{u}\hat{o}\hat{i}d-h_2a$ vs. $*\hat{u}\hat{o}\hat{i}d-h_2e$ 'I know'), vocalisation of sonorants ($*\hat{u}\hat{l}k^w\hat{o}$ - vs. $*\hat{u}\hat{l}k^w\hat{o}$ - masc. 'wolf'), the notation of Proto-Indo-European laryngeals ($*h_1 h_2 h_3$ vs. $*h, *χ, *ʰ$, in [Chapter 14](#)), dorsals ($*k g$ vs. $*q g$, again in [Chapter 14](#)), or different reconstructions of individual morphemes (e.g. $*h_2\hat{o}\hat{u}i$ - vs. $*h_3\hat{e}\hat{u}i$ - 'sheep') across chapters. I have also retained $*\hat{k} \hat{g} \hat{g}^h$ for $*\hat{k} \hat{g} \hat{g}^h$ and $*j w$ for $*\hat{j} \hat{u}$. While this practice means that the notation may differ slightly across chapters, I believe that readers who notice the discrepancies will also understand their *raisons d'être* without being confused.

It should be noted that the terminology for reconstructed stages of the Indo-European language family is not uniform. For instance, while the reconstructed ancestor of the family is referred to as "Proto-Indo-European" by some authors, others prefer "Proto-Indo-Anatolian". The terminology of the individual authors has been retained. For a discussion of the terminology describing the nodes of the Indo-European family tree, see [Olander 2019b](#).

References

- Birnbaum, Henrik & Jaan Puhvel (eds.). 1966. *Ancient Indo-European dialects*. Berkeley: University of California Press.
- Bouckaert, Remco, et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337. 957–60.
- Bouckaert, Remco, et al. 2013. Corrections and clarifications. *Science* 342. 1446.
- Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91. 194–244.
- Clackson, James. 2015. Subgrouping in the Sabellian branch of Indo-European. *Transactions of the Philological Society* 113. 4–37.

- Eichner, Heiner. 1988. Sprachwandel und Rekonstruktion: Prinzipielles zur indogermanistischen Rekonstruktion. In Christian Zinko (ed.), *Akten der 13. Österreichischen Linguistentagung, Graz 25.–27. Oktober 1985*, 10–40. Graz: Leykam.
- Garrett, Andrew. 1999. A new model of Indo-European subgrouping and dispersal. In Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the twenty-fifth annual meeting of the Berkeley Linguistics Society, February 12–15*, 146–56. Berkeley, CA: Berkeley Linguistics Society.
- Garrett, Andrew. 2006. Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 139–51. Cambridge: McDonald Institute for Archaeological Research.
- Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–9.
- Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2021. Bayesian phylolinguistics. In Richard D. Janda, Brian D. Joseph & Barbara S. Vance (eds.), *The handbook of historical linguistics*. Vol. 2, 226–53. Hoboken, NJ: John Wiley & Sons.
- Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- Kloekhorst, Alwin & Tijmen Pronk (eds.). 2019. *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*. Leiden: Brill.
- Meillet, Antoine. 1908. *Les dialectes indo-européens*. Nouveau tirage, 1922. Paris: Honoré Champion.
- Melchert, H. Craig. In press. The position of Anatolian. In *Handbook of Indo-European studies*. Oxford: Oxford University Press.
- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81. 382–420.
- Oettinger, Norbert. 2014. Die Indo-Hittite-Hypothese aus heutiger Sicht. *Münchener Studien zur Sprachwissenschaft* 67. 149–76.
- Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. Leiden: Brill.
- Olander, Thomas. 2018. Connecting the dots: The Indo-European family tree as a heuristic device. In David Goldstein, Stephanie Jamison & Brent Vine (eds.), *Proceedings of the 29th UCLA Indo-European Conference*, 181–202. Bremen: Hempen.
- Olander, Thomas. 2019a. Indo-European cladistic nomenclature. *Indogermanische Forschungen* 124. 231–44.
- Olander, Thomas. 2019b. The Indo-European homeland: Introducing the problem. In Birgit Anette Olsen, Thomas Olander & Kristian Kristiansen (eds.), *Tracing the Indo-Europeans: New evidence from archaeology and historical linguistics*, 7–34. Oxford: Oxbow.
- Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. Heidelberg: Winter.
- Ringe, Don. 1998. Some consequences of a new proposal for subgrouping the IE family. In Benjamin K. Bergen, Madelaine C. Plauché & Ashlee C. Bailey (eds.), *Proceedings of the twenty-fourth annual meeting of the Berkeley Linguistics Society, February 14–16, 1998. Special session on Indo-European subgrouping and internal relations*, 32–46. Berkeley, CA: Berkeley Linguistics Society.

- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Ross, Malcolm. 1997. Social networks and kinds of speech-community event. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language*. Vol. 1. *Theoretical and methodological orientations*, 209–61. London: Routledge.
- Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Monatsschrift für Wissenschaft und Literatur* 3. 786–7.
- Schleicher, August. 1861. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Vol. 1. *Kurzer Abriss einer Laut- und Formenlehre der indogermanischen Ursprache* [...]. Weimar: Böhlau.
- Søborg, Tobias Mosbæk. 2020. *Sigmatic verbal formations in Anatolian and Indo-European: A cladistic study*. PhD thesis, University of Copenhagen.
- Tremblay, Xavier. 2005. Grammaire comparée et grammaire historique: Quelle réalité est reconstruite par la grammaire comparée? In Gérard Fussman et al. (eds.), *Āryas, Aryens et Iraniens en Asie Centrale*, 21–195. Paris: Collège de France.

2 Methodology in Linguistic Subgrouping

James Clackson

2.1 Introduction

If two or more languages form a subgroup of a language family, what does it mean? To answer this, it will be helpful to consider the case of three related languages, A, B and C. I shall assume that these three languages are all spoken at the same point in time and are all derived from an unattested proto-language, which I shall call Proto-ABC (I shall also refer to the language family as ABC). If the languages A and B form a subgroup within ABC, this means that it is possible to reconstruct a stage intermediate between Proto-ABC and languages A and B, which I shall call Proto-AB. To put this in other words, there existed a community of Proto-AB speakers at the time when a separate speech community spoke Proto-C, the language ancestral to C. The situation can thus be represented as in [Figure 2.1](#), where languages are placed in a relationship to one another, much as with a family tree of genealogical descent.¹ Diagrams such as [Figure 2.1](#) are accordingly called “tree diagrams”.

2.2 The History of Subgrouping

The recognition of subgroups of the Indo-European language family precedes the recognition of the language family itself. [Scaliger \(1610\)](#) was already able to recognise the Romance, Germanic and Slavic families of languages, *matrices linguarum* in his terms, from shared vocabulary (notoriously using the word for ‘god’ as a diagnostic), and earlier scholars had grouped several languages as one in order to preserve the Biblical notion of seventy-two languages of the world.² From the beginning of the nineteenth century, the first scholars of Indo-European operated with subgroups such as Germanic and Slavonic. Thomas Young, in the same article which saw the first use of the term “Indo-European” arranged the languages of the world into a three-step hierarchy: classes (of

¹ [Hoenigswald \(1966: 3–5\)](#) discusses more complicated arrangements between three putative languages A, B and C.

² [Borst \(1957–63\)](#) shows in detail the changing conceptions of languages and language families in the pre-Modern era. For the background to Scaliger’s work, see [Simone \(1998: 163–5\)](#).

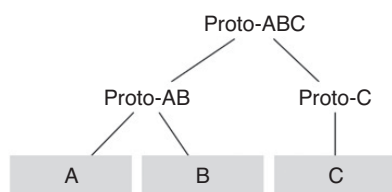


Figure 2.1 Family tree of the ABC language family

which Indo-European was one), orders and families (1813: 256). Young's Indo-European class comprised no subordinate orders, but sixteen "families", some of which are familiar, German, Celtic, Latin and Slavic, but others less so (Arabian, Etruscan and Cantabrian).³ The first representation of the relationship between languages of the Indo-European family by something like a tree diagram is generally attributed to Schleicher, who included a schematic Stammbaum at the beginning of his *Compendium* (1861: 7) although there was no explanation of how the groupings had been arrived at.⁴ The figure from Schleicher's compendium is reproduced as Figure 2.2. Unlike the diagram given in Figure 2.1, Schleicher's tree is presented with the parent language on the left, and the daughter languages on the right. The "branches" of the tree are labelled, rather than the nodes as in Figure 2.1.

The first Indo-Europeanists to give serious consideration to the methodology of language subgrouping were the "neogrammarians" (or "Junggrammatiker"), a group of scholars originally based around the University of Leipzig in the 1870s.⁵ The neogrammarians are associated today principally with the idea that sound change is regular and exceptionless, but their work on sound change was part of a larger programme which established a firmer basis for comparative linguistics. The neogrammarians were more explicit about how and why they did what they did than their predecessors, with publications on the techniques and practices of linguistic comparison.⁶ In the case of subgrouping, the first tangible advance made by the neogrammarians was Hübschmann's demonstration that that Armenian was not an Iranian language, but a separate branch on its own.⁷

³ Compare Max Müller's later division of Indo-European languages into divisions, classes and branches (Müller 1861: 380, discussed by Petit 2012: 25–7). The term Cantabrian is an alternative name for Basque, as used by Adelung.

⁴ See Petit (2012: 22–5) for discussion of an earlier tree-diagram than Schleicher's, by František Ladislav Čelakovský representing the relations between Slavic languages; Blažek (2007) gives a survey of the development of tree diagrams after Schleicher.

⁵ See Morpurgo Davies (1992: 226–78) for the neogrammarian school and its impact on linguistics.

⁶ See for example the two books on theoretical linguistics published in 1880, Delbrück 1880 and Paul 1880, discussed by Morpurgo Davies (1992: 245–51).

⁷ Hübschmann 1875; see the discussion of Hübschmann's achievement in Clackson 2016.

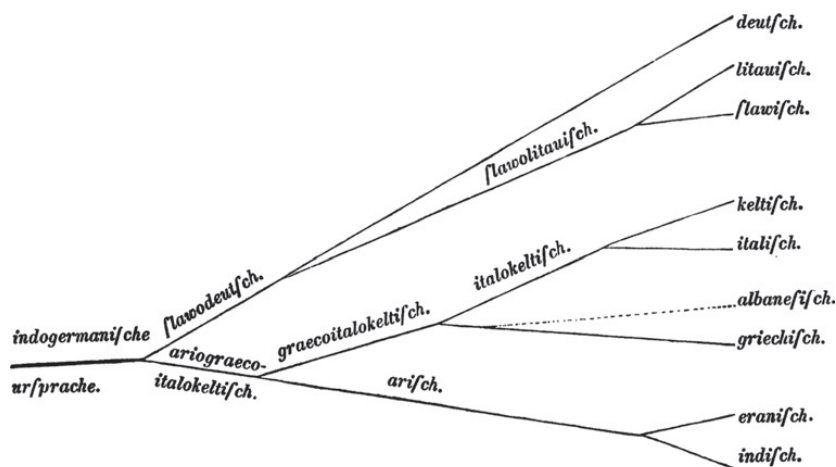


Figure 2.2 Schleicher's tree diagram (Schleicher 1861: 7).

The theoretical principles and methods set out for identifying subgroups were put forward by Leskien (1876), partly as a critical response to the Berlin professor Johannes Schmidt's work on the "wave model" (Schmidt 1872). Leskien was teacher and mentor to many of the neogrammarians, and his work on subgrouping was then refined by Delbrück (1880: 135) and Brugmann (1884).⁸ The methodological advances made by these scholars are enormous. It is to them that we owe the principles that linguistic subgrouping proceeds through the identification of shared innovations, rather than shared archaisms, and the recognition that phenomena which could arise from language contact, such as shared lexical items, should be treated with caution for subgrouping purposes.⁹

Indeed, Brugmann's statement of what constitutes a subgroup (1884: 253) has often been cited, and is worth repeating once again:

Es ist hier nicht eine einzelne und sind nicht einige wenige auf zweien oder mehreren Gebieten zugleich auftretende Spracherscheinungen, die den Beweis der näheren Gemeinschaft erbringen, sondern nur die große Masse von Übereinstimmungen in lautlichen, flexivischen, syntaktischen und lexicalischen Neuerungen, die große Masse, die den Gedanken an Zufall ausschließt.

⁸ In Morpurgo Davies's words "the neogrammarians, as often, took their cue [*sic*] from Leskien" (1975: 650).

⁹ Morpurgo Davies (1975: 650) and Petit (2012: 29–30) associate these ideas directly with Leskien, but as I showed in a recent paper (Clackson 2016), they are already implicit in Hübschmann's (1875) work on Armenian.

[The proof of a close commonality comes not from a single isolated or a small number of linguistic developments occurring simultaneously in two or more areas, but only through a large number of innovations in phonology, morphology, syntax and vocabulary – a number so great as to exclude chance from consideration.]

Brugmann's 1884 article has set the scene for the subgrouping of the Indo-European family and other language groupings ever since.¹⁰ Brugmann discussed the possible subgroups of Indo-European, seeing only two cases where the recognised branches of Indo-European might be grouped together: Indo-Iranian and Balto-Slavic. It is significant that since 1884 there have been no serious suggestions for some of the higher order groupings proposed seen in Schleicher's family tree, and the Indo-European family continues to be thought of in terms of the branches Brugmann identified.¹¹ After the neogrammarians Schleicher's "Graecoitalokeltisch" and "Slawodeutsch" all but disappear from the academic debate for the next hundred years.¹² Representations of the Indo-European family in tree diagrams in the century after Brugmann's article tended to show the branches of Indo-European radiating out as spokes from a centre.¹³ Indeed, the discovery of two new branches of Indo-European in the early twentieth century, Anatolian (of which Hittite was the earliest identified) and Tocharian, had little initial impact on the presentation of the Indo-European languages. Bloomfield's tree diagram (Bloomfield 1933: 315) does not include branches for Anatolian or Tocharian, and Meillet was able to issue a second edition of a book written originally in 1908 (and discussed further below) in 1922 only noting the recent addition of the two branches (Meillet 1922: 3).

2.3 Criteria for Subgrouping

The reliance on common innovations rather than common retentions, and the need to avoid linguistic agreements that could have arisen independently, or by chance, have been accepted by nearly all those working on subgrouping methodology since Brugmann.¹⁴ It has been suggested (Dyen 1953: 581–2) that, despite linguists' theoretical adherence to the methodology of Brugmann,

¹⁰ See the discussion of Dyen (1953: 580–2), who is the first to use the term "subgrouping" in English.

¹¹ See for example the presentation of the Indo-European languages in Fortson 2010 and Klein, Joseph & Fritz 2017–18.

¹² A Balto-Slavic-Germanic subgroup reappears in the tree-diagram of Gamkrelidze & Ivanov (1984: 415).

¹³ To give just two examples, Bloomfield 1933: 312 and the representation of the Indo-European language family in editions of the *American heritage dictionary of the English language* (first published in 1969).

¹⁴ See Porzig 1954: 17–52 and Clackson 1994: 4–11 for surveys of work on Indo-European subgrouping in the twentieth century; Ringe & Eska (2013: 256–7) and Ringe (2017: 63) have recently reiterated the need to base subgroups on significant shared innovations.

most subgrouping has actually been carried out “by inspection”, that is to say, through the recognition of a large amount of similarities between closely related languages (much as Scaliger was able to recognise that the Romance languages or the Germanic languages belonged together). This may be true at a very basic level, but any serious considerations of subgrouping for individual languages since the 1870s have proceeded through careful application of something like the Brugmannian criteria. This is especially the case for the less well-attested Indo-European varieties, such as Phrygian, Venetic or Lusitanian. If scholars have not used Brugmann’s criteria to test the validity of the Germanic branch, or Slavic, it is because the innovations are numerous and self-evident.

In the rest of this chapter, I shall look first at further clarifications of the criteria for subgrouping given above, before considering alternative models to the family tree. Advances in the neogrammarian methodology outlined above have been made in three principal areas: assessment of what counts as an innovation; ways to avoid “false positives”, that is, apparent shared innovations which actually arise by chance or through language contact; and in the use of computational methods in order to survey large amounts of data (see [Chapters 3 and 4](#)).¹⁵ I shall discuss the first two of these developments in this chapter, leaving the third to other contributors to this volume.

How do linguists recognise an innovation against a shared retention? In some areas, innovations are easier to detect than others. Once speakers of a language have merged or partially merged two phonemes, this change cannot be undone. Consequently, phonological changes offer the clearest examples of innovations which can be recovered by the historical linguist. As Hoenigswald put it (1966: 7), the phonological merger is the “prototype” of the shared innovation. Vocabulary replacement and syntactic change provide examples where it is often more difficult to isolate which development is an innovation and which is not. If languages A and B share a vocabulary item, for example the word for ‘man’ or a verb used to mean ‘stand’, and this vocabulary item is not found in language C, how is it possible to ascertain whether that is a shared retention or a common development of the two languages? Dyen is one of only a few scholars to address this question directly:

If any two or more related languages share a feature, the question arises whether this is a retention or an innovation. If we apply a general rule that such features are taken to be retentions unless there is evidence to the contrary, then a corresponding proto-feature is reconstructed. It follows that (borrowings being excluded) an innovation occurring in

¹⁵ See [Ringe, Warnow & Taylor 2002](#): 66 for the statement that the computational approach to subgrouping “is *not* intended to replace already existing methods, but to supplement them” (emphasis in the original).

two or more languages can be detected only if, as a proto-feature, it contradicts a proto-feature which for some reason appears to be more ancient. (Dyen 1953: 581)

In the Indo-European domain, researchers have the advantage that the Anatolian languages, now generally agreed to have split first from the Proto-Indo-European parent, can sometimes provide a guide to what forms are more “ancient”. To take one example, Greek, Armenian, Albanian and Tocharian share reflexes of a root ‘hand’, reconstructed as **ǵ^{hes}-r-* (Greek *χεῖρ*, Armenian *jern*, Albanian *dorë*, Tocharian A *tsar*). For Pedersen (1924: 225), followed by Solta (1960: 316–17), this was a significant lexical agreement between these branches. However, it is now clear that the word is also present in the Anatolian branch; the presence of the word in the other languages is much more likely to be a shared retention rather than an innovation.¹⁶ Vocabulary items may also be judged to be archaic rather than innovatory through their inflectional class or derivational patterns, or because it is possible to reconstruct a semantic shift in one direction rather than another. Even so, such decisions are often reliant on the judgement of the linguist, and in many cases it is impossible to say whether a lexical agreement reflects an innovation or a retention (Hoenigswald 1966: 8–9; Klingenschmitt 1994: 236).

Innovations in inflectional morphology are also to some extent reliant on the picture the researcher has of the morphology of the parent language, and hence susceptible to the same criticism as the use of shared items of vocabulary for subgrouping purposes. Morphological innovations are, however, generally easier for linguists to spot, since they may be linked to phonological changes and thus more easily linked to a relative chronology.¹⁷ Moreover, in inflectional morphology at least, the set of options which can be reconstructed for the parent language is in general much smaller than for lexical innovations. Morphological innovations may also be associated with a larger change in the morphosyntax of the language, such as the creation of a new category or the merger of earlier categories.

In the Indo-European language family, little use has been made of syntactic changes for subgrouping, partly reflecting uncertainties about the reconstruction of Indo-European syntax, with consequent uncertainty about what counts as an innovation.¹⁸ In this regard it is important to note recent attempts to find Indo-European subgroups relying on syntactic information put forward by Longobardi & Guardino (2009) and Longobardi et al. (2013). These researchers, working in a Chomskyan syntactic framework, make use of a set

¹⁶ Hittite *kessar*, Hieroglyphic Luwian *istra/i-*, Lycian *izre/i-* (see Klockhorst 2008: 471–2); for the use of the word for ‘hand’ in a recent subgrouping enterprise, see Ringe, Warnow & Taylor 2002: 82–3.

¹⁷ See further below for the importance of relative chronology.

¹⁸ Ringe, Warnow and Taylor (2002) include no syntactic features in their data set.

of syntactic parameters, which have been carefully chosen to ensure that the selected parameters show no overlap between them. It is perhaps revealing that the researchers do not attempt to isolate which of the parametric changes count as innovations, relying on the computational approach to identify innovations (Longobardi et al. 2013: 148). Given the absence of suitable information about the parametric constraints of most of the older Indo-European languages, this approach has not proved to be especially helpful for refining current thinking on subgrouping in the language family.

The next problem is how to avoid “false positives”, that is, shared innovations between two languages which were not made during a period of genealogical unity but which come about at a later stage in the language histories. In terms of the hypothetical language family discussed at the beginning of this chapter, examples would be developments shared by languages B and C that took place after the break-up of the Proto-ABC community, or shared by A and B but made after the period of Proto-AB. Such shared developments among separate speech communities may reflect a situation of language contact, for example a period when many speakers of A also spoke B, or when speakers of A and B both spoke a third language, or another more complicated contact situation. Alternatively, a shared development made independently is sometimes attributed to “chance”. In effect, what this usually means is that the innovation may reflect a universal tendency of language development, such as the palatalisation of dorsal consonants before front vowels or the “drift” from perfect formations of the verb to perfectives.¹⁹ Languages of the same family have inherited similar structures, and it is consequently not unexpected that the same innovations may occur independently.

As has been recognised since Meillet (1908: 10), understanding the relative chronology of changes is essential in order to determine which shared developments are common shared innovations and which are not. In the terms of the ABC language family, an innovation which is apparently shared by A and B is not diagnostic for subgrouping if it can be shown to have taken place after a development that took place after A had split from B. To take an example from the Italic language family, Oscan and Umbrian have both undergone a process of syncope of short medial vowels so that, for example, an earlier stem **opesā-* develops to an Oscan stem *úpsa-* and Umbrian *osa-*. But this change is fed by consonant changes in Umbrian, such as the development of intervocalic **d > rs* and the palatalisation of velars before front vowels. Syncope, which is not an uncommon change cross-linguistically, is thus an innovation shared by Oscan and Umbrian but is not diagnostic for their

¹⁹ For a survey of the sound changes affecting consonants which occur across Indo-European languages, see Kümmel 2007; on universal paths in the grammaticalisation of tense and aspect, see Bybee, Perkins & Pagliuca 1994, which has spawned a large body of work on processes such as the drift from perfects to perfectives (sometimes called “aoristic drift”).

subgrouping, since it must have taken place after Umbrian changes not shared by Oscan (see [Clackson 2015](#): 10).

In the search for diagnostic innovations for subgrouping, it may not always be possible to construct a relative chronology for the feature in question in relation to what else is known about the prehistory of the language family. Accordingly, linguists have looked to assess the likelihood that a particular innovation might be the result of contact or universal processes, rather than a shared innovation. The result is that some shared innovations carry more “weight” than others, which may be dismissed as “trivial” or “insignificant”.²⁰ Phonological processes which are frequent across the languages of the world, such as palatalisation, lenition, apocope, are accordingly usually dismissed as easily replicable and non-diagnostic. Many scholars have given greater weight to less common or more “unusual” sound changes, although in the absence of a general cross-linguistic repertoire of all known sound changes, this may rely more on the researcher’s own knowledge than an objective assessment. Note also that in the Indo-European language family, the judgement of whether a change affecting reconstructed consonants such as “laryngeals” or “voiced aspirates” is unusual or not also reflects the reconstructed model of PIE which is used. Individual shared vocabulary which might arise from borrowing from languages now lost is similarly easily discounted for subgrouping purposes. Once again it is innovations in the field of morphology, particularly inflectional morphology, which has been seen as especially significant. Incorporation of the inflectional morphology of one language into another is not unknown in situations of prolonged or close contact, or in particular social situations, but it is generally accounted the most resistant area of language to borrowing.²¹ The creation of a new morpheme often reflects the grammaticalisation of a new category or the merger of earlier categories, and accordingly morphological innovations generally have significant structural importance in the languages in question.

The question remains of how many innovations is enough to reconstruct a subgroup? Brugmann rejected the reliance on a “few” innovations, calling rather for a “large number”, but then he had not made the various further refinements of sorting through what were certain, appropriate or significant innovations, using the methodology developed by later scholars and outlined above. Once all potential shared innovations between two branches have been

²⁰ The “weighting” of isoglosses is implicit already in [Hübischmann 1875](#), and highlighted by [Meillet \(1908\)](#) and [Porzig \(1954\)](#).

²¹ See [Thomason & Kaufman 1988](#): 18–20 for the dismissal of earlier claims that morphology is impervious to borrowing. Morphological borrowing is not just limited to “exotic” languages: the Latin first declension genitive *-aes*, found predominantly in texts written by writers with little education, shows the partial transfer of a Greek morpheme (see [Adams 2003](#): 473–86 for discussion).

carefully sifted to determine whether they meet all suitable criteria, and those for which there remains any room for doubt have been set to one side, then is it still justified to say that the number remaining is too small to be significant?²² Recent family trees of Indo-European arrived at using computational cladistics, which may have examined a large set of data across vocabulary, morphology and phonological changes (as Ringe, Warnow & Taylor 2002), show a much greater number of branches and subgroups than most of those constructed following Brugmann's 1884 article.²³ In the Ringe, Warnow and Taylor tree, binary splits are the norm, as opposed to the fan-like array of earlier trees. The difference partly reflects the ways in which the computational analysis is constructed, but it also reflects the fact that some of the subgroups are constructed on what is in effect quite a small number of shared features. The Greco-Armenian clade, for example, is supported by only six shared lexical features, four of which need not be significant.²⁴

2.4 Subgroups and Prehistoric Dialect Continua

So far in this chapter, I have largely followed the assumption that language change operates over uniform speech communities and that language diversification happens when a single speech community splits into two or more separate groups. However, linguistic history is rarely so straightforward. Clean breaks in the tree-diagram, such as that envisaged in our opening example between Proto-AB and Proto-C, may occur as the result of large-scale dispersals of a population after cataclysmic natural disasters, through massive migrations or other situations, but in the majority of documented situations, the diversification of a language into separate, mutually unintelligible, descendants takes place through periods of dialect continua, which might sometimes last for millennia. Indeed, the spoken varieties of Romance, Germanic, Slavic and several other branches of the Indo-European family still can be described, in whole or in part, as dialect continua. Since Schmidt (1872), linguists have recognised that the spread of phenomena over dialect continua are not best captured by a tree-diagram model. Schmidt himself famously proposed an alternative to the tree diagram, the “wave theory” (“Wellentheorie”), to explain the rippling effect of linguistic changes over a range of mutually comprehensible varieties.²⁵

Leskien and the neogrammarians made a significant advance on Schmidt's observations by pushing the period of dialectal variation back

²² This is a criticism that has been levelled at me (see, for example, Holst 2009: 53–5) for my “hyper-critical” analysis of the evidence for a Greek-Armenian subgroup (Clackson 1994).

²³ See the survey in Blažek 2007.

²⁴ As noted by Ringe, Warnow & Taylor (2002: 102–3), Ringe (2017: 69).

²⁵ See the discussion by Petit (2012: 27–9).

to the proto-language, rather than, as Schmidt had suggested, a period when it was possible to recognise the first branches of Indo-European. Their methodological justification for this move was that all spoken languages contain some variation, and it is consequently likely that the proto-languages also exhibited variation.²⁶ This line of reasoning was followed up by scholars in the early twentieth century, such as Meillet, whose 1908 book *Les dialectes indo-européens* explored at greater length various shared developments of vocabulary, phonology and morphology that might reflect dialectal divisions within the parent language (Meillet 1908, second edition 1922). For example, the noteworthy shared agreement of Germanic, Baltic and Slavic in showing **m* rather than **b^h* in oblique case markers of the noun could only be explained, according to Meillet, through the supposition of different dialects of Proto-Indo-European (1908: 119).²⁷ The reconstruction of dialects of the proto-language thus allowed historical linguists a way to account for a small number of similarities between languages which were not sufficient on their own to support the reconstruction of a subgroup but were too significant to be ignored. As we have seen, the net effect of this move was that, in contrast to the recognised subgroups lower down the family tree, such as Germanic, Celtic etc., after Brugmann (1884), there were only two generally agreed “higher-order” subgroups, Indo-Iranian and Balto-Slavic. The supposition of a “dialectal” Proto-Indo-European could help explain the existence of a small number of exclusive and significant innovations shared between two or more branches, and also the overlapping nature of these agreements, so that some features might be shared between Germanic and Balto-Slavic, and others between Balto-Slavic and Indo-Iranian.

The supposition of a dialectal array of Indo-European has consequently proved popular and in several handbooks of historical linguistics or Indo-European languages it is possible to find “dialect maps” of Proto-Indo-European, with the putative varieties ancestral to the different branches of the family laid out in something approximating to their geographical attestation, with Germanic at the top left and Indo-Iranian in the bottom right, and then isogloss lines linking or separating groups corresponding to shared “dialectal” features, such as the use of **m* or **b^h* in oblique case-markers or the operation of a phonological process known as the *ruki* rule.²⁸ Such maps meet with the immediate criticism that they

²⁶ As noted by Petit (2012: 31) who cites Leskien (1876: xv): “auf dem Boden der Urheimat [bestanden] bereits dialektische Unterschiede” [“there were already dialectal differences in the territory of the (Indo-European) homeland”].

²⁷ Bloomfield (1933: 314–5) also uses the example of the **m* and **b^h* case markers as indications of dialectal differences in PIE.

²⁸ Meillet’s schematic map (1908: 134) has been followed by many others. Anttila 1989: 305 is the most sophisticated with twenty-four isoglosses included; Hock 1991: 445 has seven isoglosses, Mallory & Adams 2006: 73 just six.

have the potential to include items of different time depths on a single plane. Thus, if the Anatolian branch separated out from the other IE languages first, it is an anachronism to include it in a dialectal area which could not yet have existed.²⁹ Meid (1975) has accordingly attempted to reconstruct a “space-time” model for PIE, which is in effect a “three-dimensional” dialect map, incorporating both temporal and dialectal variety.

The reconstruction of a Proto-Indo-European parent language which has variation over time and space meets with a significant methodological objection: it is difficult to falsify. As Ringe (2017: 65) elegantly expresses it: “new evidence that is at variance with evidence already in hand can often be accommodated on an abstract dialect ‘map’ without major revisions.” Moreover, it significantly overplays the importance of the evidence which happens to survive. Since 1950 the number of early Indo-European texts available to scholars has been significantly increased through greater knowledge of the Anatolian languages; the decipherment of Linear B and consequent accessibility of the earliest stage of Greek; and discoveries and improvements in the understanding of a number of smaller, fragmentary languages, such as Gaulish, Celtiberian and South Picene. This huge increase in our knowledge of the languages used in the first and second millennia BCE has, paradoxically, made scholars more aware of what has been lost. It is clear that, since the Iron Age, speakers of a relatively small number of language families and subfamilies have been hugely successful in Eurasia, and their dominance has been responsible for the demise of countless other languages, many of which were Indo-European. As Ringe & Eska (2013: 262–3) note, the branches of Indo-European that we know about are “probably the surviving remnants of what was once a dialect network”, and the apparent sharp distinctions between them are just the reflection of the “pruning” of closer neighbours. Garrett has suggested in a number of articles that this loss of the intermediate languages in a larger dialectal continuum means the dialectal array of Proto-Indo-European after the separation of Anatolian, and probably Tocharian, may not be retrievable (see Garrett 1999; 2006; Babel et al. 2013). This is not just because of the pruning problem but also because of the fact that the subgroups as we have them may reflect linguistic changes across a dialect continuum, which took place across already divergent dialects. A case in point is Greek, which shows dialectal divisions already in the Mycenaean period, but for which all dialects were to undergo significant shared innovations in the next 500 years. These subsequent wave-like innovations take on a special significance when we have lost so many of the intervening dialects. The combination of

²⁹ In the map of Anttila (1989: 305), Anatolian sits in the middle separated from all other languages by two isoglosses, one of which is drawn with a thicker line. The maps of Hock (1991: 445) and Mallory & Adams (2006: 73) do not include Anatolian.

shared innovation across an already differentiated dialectal continuum and subsequent “pruning” of intermediate dialects means that the shape of the original dialectal array is forever unobtainable.

Few Indo-Europeanists have been willing to accept Garrett’s arguments for scepticism about subgrouping, however, and most have continued to operate with a branching tree model, with shared innovations as the diagnostic for the construction of a subgroup.³⁰ The objections to Garrett’s proposals are founded on a reluctance to give weight to the “unknown unknowns”, that is the unrecorded Indo-European varieties which gave way to the languages which we know about, and which may have formed a bridge between what we now think of as different Indo-European subgroups (see the comment recorded by Garrett of an anonymous referee at 2006: 48 n. 5; de Vaan 2008: 1229–30). These varieties doubtless existed, but we don’t know how they would have changed our whole picture of Proto-Indo-European as a whole or, indeed, what they would have been like. To abandon the whole enterprise of subgrouping because we don’t know what we are missing seems a step too far. Moreover, there has been no conclusive demonstration of an Indo-European subgroup that has actually arisen through later convergence.³¹ It is likely that subgrouping as currently carried out will continue, even though Garrett’s arguments are a healthy reminder of the importance of considering the relative chronology of linguistic developments, and of guarding against the false reconstruction of a subgroup on the basis of changes which must have actually been convergences.

References

- Adams, J. N. 2003. *Bilingualism and the Latin language*, Cambridge: Cambridge University Press.
- Anttila, Raimo. 1989. *Historical and comparative linguistics*. 2nd rev. ed. Amsterdam: Benjamins.
- Babel, Molly, Andrew Garrett, Michael J. Houser & Maziar Toosarvandani. 2013. Descent and diffusion in language diversification: A study in Western Nomic dialectology. *International Journal of American Linguistics* 79. 445–89.
- Blážek, Václav. 2007. From August Schleicher to Sergei Starostin: On the development of the tree-diagram models of the Indo-European languages. *Journal of Indo-European Studies* 35. 82–110.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston.

³⁰ Clackson 2007: 14–15 is an exception, but reviewers have taken issue with this point, for example de Vaan (2008). See also Goldstein 2020: 113, who reiterates the point that the true tree for the PIE phylogeny is unknowable.

³¹ My attempt to explain similarities between languages in the Sabellian subgroup, including Oscan and Umbrian, through convergence (Clackson 2015) has been countered by Fortson (2017).

- Borst, Arno. 1957–63. *Der Turmbau von Babel: Geschichte der Meinungen über Ursprung und Vielfalt der Sprachen und Völker*. 5 vols. Stuttgart: Hiersemann.
- Brugmann, Karl. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen. *Internationale Zeitschrift für allgemeine Sprachwissenschaft* 1. 226–56.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago, IL: University of Chicago Press.
- Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.
- Clackson, James. 2007. *Indo-European linguistics*. Cambridge: Cambridge University Press.
- Clackson, James. 2015. Subgrouping in the Sabellian branch of Indo-European. *Transactions of the Philological Society* 113. 4–37.
- Clackson, James. 2016. Heinrich Hübschmann and linguistic classification. In Jean-Léo Léonard (ed.), *Actualité des néogrammairiens*, 183–201. Paris: Peeters.
- Delbrück, Berthold. 1880. *Einleitung in das Sprachstudium: Ein Beitrag zur Geschichte und Methodik der vergleichenden Sprachforschung*. Leipzig: Breitkopf und Härtel.
- Dyen, Isidore. 1953. Review of Otto Chr. Dahl, *Malgache et Maanjan: Une comparaison linguistique*, Oslo: Gimnes, 1951. *Language* 29. 577–90.
- Fortson, Benjamin W., IV. 2010. *Indo-European language and culture: An introduction*. 2nd ed. Malden, MA: Wiley-Blackwell.
- Fortson, Benjamin W., IV. 2017. The dialectology of Italic. In Klein, Joseph & Fritz 2017–18, vol. 2, 835–57.
- Gamkrelidze, T. V. & V. V. Ivanov 1984. *Indoevropskij jazyk i indoevropejcy*, Tbilisi: Tbilisi University Press.
- Garrett, Andrew. 1999. A new model of Indo-European subgrouping and dispersal. In Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the twenty-fifth annual meeting of the Berkeley Linguistics Society, February 12–15*, 146–56. Berkeley, CA: Berkeley Linguistics Society.
- Garrett, Andrew. 2006. Convergence in the formation of Indo-European sub-groups: Phylogeny and chronology. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 139–51, Cambridge: McDonald Institute for Archaeological Research.
- Goldstein, David. 2020. Indo-European phylogenetics with R: A tutorial introduction. *Indo-European Linguistics* 8. 110–80.
- Hock, Hans Henrich. 1991. *Principles of historical linguistics*. 2nd rev. and updated ed. Berlin: Mouton de Gruyter.
- Hoenigswald, Henry M. 1966. Criteria for the sub-grouping of languages. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 1–12. Berkeley: University of California Press.
- Holst, Jan Henrik. 2009. *Armenische Studien*. Wiesbaden: Harrassowitz.
- Hübschmann, H. 1875. Ueber die stellung des armenischen im kreise der indogermanischen sprachen. *Zeitschrift für vergleichende Sprachforschung* 23. 5–49.
- Klein, Jared S., Brian D. Joseph & Matthias Fritz (eds.). 2017–18. *Handbook of comparative and historical Indo-European linguistics*. 3 vols. Berlin: De Gruyter Mouton.

- Klingenschmitt, Gert. 1994. Die Verwandtschaftsverhältnisse der indogermanischen Sprachen. In Jens Elmegård Rasmussen (ed.), *In honorem Holger Pedersen*, 235–51. Wiesbaden: Reichert.
- Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.
- Leskien, August. 1876. *Die Declination im Slavisch-Litauischen und Germanischen*. Leipzig: Fürstlich Jablonowskische Gesellschaft der Wissenschaften.
- Longobardi, Giuseppe & Cristina Guardino. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119. 1679–1706.
- Longobardi, Giuseppe, et al. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3. 122–52.
- Mallory, James P. & Douglas Q. Adams. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford: Oxford University Press.
- Meid, Wolfgang. 1975. Probleme der räumlichen und zeitlichen Gliederung des Indogermanischen. In Helmut Rix (ed.), *Flexion und Wortbildung*, 204–19. Wiesbaden: Reichert.
- Meillet, Antoine. 1908. *Les dialectes indo-européens*. Paris: Champion.
- Meillet, Antoine. 1922. *Les dialectes indo-européens*. 2nd ed. Paris: Champion.
- Morpurgo Davies, Anna. 1975. Language classification in the nineteenth century. In Thomas Sebeok (ed.), *Current trends in linguistics* 13, 607–716. The Hague: de Gruyter.
- Morpurgo Davies, Anna. 1992. *Nineteenth-century linguistics*. London: Longman.
- Müller, Max. 1861. *Lectures on the science of language delivered at the Royal Institution of Great Britain in April, May, and June, 1861*. London: Longman, Green, Longman, and Roberts.
- Paul, Hermann. 1880. *Principien der Sprachgeschichte*. Halle: Niemeyer.
- Pedersen, Holger. 1924. Armenier. In Max Ebert (ed.), *Reallexikon der Vorgeschichte*. Vol. 1, 219–26. Berlin: de Gruyter.
- Petit, Daniel. 2012. Débats sur la classification des langues indo-européennes au xix^e siècle. *Lalies* 32. 9–32.
- Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. Heidelberg: Winter.
- Ringe, Don. 2017. Indo-European dialectology. In Klein, Joseph & Fritz 2017–18, vol. 1, 62–75.
- Ringe, Don Joseph F. Eska. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Scaliger, Joseph Justus. 1610. Diatriba de europaeorum linguis. In *Opuscula varia antehac non edita*, 115–18. Paris: Drovart.
- Schleicher, August. 1861. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: Böhlau.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.

- Solta, Georg Renatus. 1960. *Die Stellung des Armenischen im Kreise der indogermanischen Sprachen*. Vienna: Mekhitaristen.
- Simone, Raffaele. 1998. The Early Modern period. In Giulio Lepschy (ed.), *History of linguistics*. Vol. 3. *Renaissance and Early Modern linguistics*, 149–236. London: Longman.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- de Vaan, Michiel. 2008. Review of Clackson 2007. *Lingua* 118. 1228–32.
- Young, Thomas. 1813. Review of Johann Christoph Adelung, *Mithridates, oder Allgemeine Sprachenkunde*, vol. 1–3, Berlin: Voss, 1806. *The Quarterly Review* 10. 250–92.

3 Computational Approaches to Linguistic Chronology and Subgrouping

Dariusz Piwowarczyk

3.1 Computational Approaches to Historical Linguistics

Computational approaches to historical linguistics can be roughly divided into those pertaining to language classification (e.g. [Ringe et al. 2002](#)), chronology (e.g. [Gray & Atkinson 2003](#)), cognate detection (e.g. [Kondrak 2002](#)), comparative reconstruction (e.g. [Hewson 1974](#)) and the simulation of phonological ([Baker 2008](#)) and analogical change ([Skousen 1989](#)).¹ They usually involve quantitative methods to calculate the relationship between the languages or to date the chronology of their split from the proto-language, as well as algorithms to align cognates, reconstruct proto-forms or simulate sound changes (including artificial neural networks for analogical changes).

Although the use of quantitative methods is not new in contemporary linguistics – they were already being used in the 1930s – the application of computational methods in historical linguistics, like those used in evolutionary biology, represents a novel approach which is gaining adherents but is mostly regarded as problematic by traditional historical linguists. This is partly because of the history of quantitative approaches, which included methods such as glottochronology and lexicostatistics that are now found to be largely unreliable (cf. [Bergsland & Vogt 1962](#)) or controversial at the very

This work has been written under the research project financed by the National Science Centre (Poland) decision number: 2018/02/X/HS2/03669. I am grateful to Thomas Olander, Don Ringe and Michael Weiss for comments on the earlier version and to Pete Westbrook for correcting my English. Needless to add, I am solely responsible for errors and mistakes.

¹ Cf. [Sims-Williams 2018](#) for a recent short overview of past approaches. Computational approaches to historical linguistics have been expanding ever since the early 2000s. Nowadays, there is an enormous outgrowth in works dealing with computational historical linguistics in all its aspects (language classification, cognate alignment, sound change simulation, analogical change simulation). For an overview of the most recent approaches, see [Dunn 2015](#) and [Jäger 2018](#). For a short overview and assessment of quantitative approaches to historical linguistics, see the discussion in the third edition of Lyle Campbell's handbook ([Campbell 2013](#): 447–92).

least (cf. [Hoiijer 1954](#)). Additionally, many linguists approached such methods with extreme caution because they involved handling and converting the data to a machine-readable format, using statistical algorithms which seemed to generate a “black-box” effect rather than explain the results, and comparing language development to the replacement of genetic material in evolutionary biology. Compounding the problem was the fact that many of these early computational approaches were actually implemented by computational biologists rather than linguists.

Even though the methods and the data used have often not been applied very methodically or carefully (especially in the early works dealing with linguistic classification and chronology), there is no doubt that computational methods can be useful in linguistics, simply because computers can analyse masses of data in a short space of time and without errors. This might not be helpful in all aspects of historical linguistics, but it will certainly make it easier to check, for example, the coherence of a hypothesis put forward, to test competing hypotheses or the results of the application of a sound change etc. However, it has to be borne in mind that:

1. The computer is only a tool and the results yielded will always ultimately depend on the quality of the algorithm, the input data and how they are converted into a machine-readable format (together with all the judgements made by the researchers at this point).
2. Research results from using computational methods have to be interpreted, and the method itself is usually meant as a supplement to traditional methods, not a replacement.

In this chapter, I will outline the history of computational approaches to historical linguistics, concentrating on those concerning language classification and chronology, and then describe the three major contrasting approaches to the classification and chronology of the Indo-European languages. Following this, I will present the method used for the computational replication of sound change, which is a useful tool for testing the relative chronology of sound changes and may have some bearing on the grouping of the Indo-European languages. I conclude with some perspectives for future work.

3.2 Sources and Origins

Although most of the contemporary computational approaches to language classification and chronology stem directly from the methods used in evolutionary biology (cf. [Dunn 2015](#)), they seem to have their indirect roots in the earlier quantitative approaches which used statistical and mathematical methods to calculate different aspects of language change and comparison.²

² Only the main approaches are outlined here. For details on the different statistical approaches and their history, I refer the reader to the overview of the use of statistics in historical linguistics by

This clearly stems from the ongoing search for more objective bases to support the traditional arguments from historical linguistics, which were often seen as subjective as they depended largely on the assumptions of the scholars who performed the research.

One of the earliest quantitative approaches was a method devised by the Polish anthropologist Jan [Czekanowski \(1927\)](#), who tried to calculate the similarities between the Indo-European languages and present the results in a numerical way. Czekanowski approached the Polish linguist Jerzy Kuryłowicz for a list of characteristic binary features which could be used in distinguishing the different subgroups of Indo-European (at first twenty, then twenty-two) and proceeded to create the two-feature contingency tables in which those features were counted for every language that was being compared. Then he used Pearson's tetrachoric correlation formula known from statistics. This allowed him to present a distance matrix of the Indo-European languages. His approach made its way to the United States through his student in anthropology, Stanisław Klimek, who came to study with Alfred Kroeber in the 1930s. Czekanowski's method was adopted by both [Alfred Kroeber and Charles Douglas Chrétien \(1937\)](#), who tried to count the similarities between the Indo-European languages using a broader range of features (seventy-four in total) taken from Meillet's monograph on the Indo-European dialects (1922). However, their results were criticised (cf. [Safarewicz 1948](#)) for being biased from the very start because of the use of data based on a work which itself intended to prove the groupings of, for example, Italo-Celtic ([Meillet 1922](#)). Probably the most famous approach was championed in the 1950s by Morris [Swadesh \(1952\)](#) in the form of glottochronology. Although promising at first, it was harshly criticised for working on the initial assumption that there is a constant rate of change in languages. [Embleton \(1986\)](#) tried to further enhance the computerised methods of lexicostatistics and glottochronology and concluded that

for the traditional methods as well as the statistical methods the reconstruction of the topology of the tree is more accurate than the assignment of dates. Reliable dating information is more likely to come from historical or archeological sources, although the statistical methods can provide some provisional estimates. ([Embleton 1986](#): 169–70)

3.3 Computational Approaches to Language Classification and Chronology

As mentioned above, most of the contemporary computational approaches to language classification and chronology stem from the methods used in

Sheila [Embleton \(1986\)](#) and the monograph on word lists and lexicostatistical approaches to language comparison by Brett [Kessler \(2001\)](#). For an overview and a summary assessment of both glottochronology and early lexicostatistics, see [Tischler \(1973\)](#).

biological sciences.³ The computational approaches used in evolutionary biology were applied to linguistics in the late 1990s and began to be used in the early 2000s. They usually use statistical Bayesian inference to infer phylogenies. This kind of work has become very popular and has already been applied to different families of languages.

An interesting distance-based approach was pioneered by Søren Wichmann and his team (2018) in the Automated Similarity Judgement Program and the corresponding ASJP database (Wichmann et al. 2018). The database includes a list of forty basic words for more than 5,000 languages and can be used, for example, to date when the languages in one family split away from each other. Because it uses the Levenshtein distance and lexical data, it is often regarded sceptically by linguists (cf. Greenhill 2011).⁴ Additionally, there are some errors in the database itself. In the word list for Latin, for example, there is no vowel length present, and the words are transcribed inconsistently: *wenire* ‘to go’ is transcribed with /w/ whereas *viya* ‘road’ is transcribed with /v/.

One of the most controversial aspects of computational (or more accurately statistical) approaches to language classification and chronology is the fact that they are heavily based (often even exclusively based) on lexical data. In contrast, the standard procedure in traditional historical linguistics is the analysis of phonological and morphological features, and this is probably the main reason that many traditional historical linguists are generally very sceptical about using computational approaches for language classification and chronology. As pointed out by Gerhard Jäger and Johann-Mattis List in their recent comparison of traditional and computational methods, the crucial difference between the classical comparative method and the approaches adopted by computational historical linguistics is that “the comparative method strives to reconstruct the *true* history of languages in their entirety while statistical approaches search for *probable* or at least *useful* models of the observed patterns in some well-defined partial range of data”.⁵

3.4 Application to Indo-European Studies

Apart from the lexicostatistical approach employed by Dyen et al. (1992) using the 200-word Swadesh lists for ninety-five languages and assuming a similar rate of change in all languages, probably the most famous computational

³ For an in-depth overview of the methods, I refer the reader to Nichols & Warnow (2008) and, for a discussion of more recent studies in the area, to Dunn 2015.

⁴ “The Levenshtein distance is a simple distance metric derived from the number of edit operations needed to transform one string into another” (Greenhill 2011: 689).

⁵ Gerhard Jäger & Johann-Mattis List, 2019, Statistical and computational elaborations of the classical comparative method (unpublished manuscript), <https://bit.ly/3yVktOs> (accessed 20 February 2020): 30.

classification of Indo-European languages was developed in 2002 by a team of experts combining linguistics (Don Ringe, Ann Taylor) and computer science (Tandy Warnow) in the project on *Computational phylogenetics in historical linguistics* (with contributions from statistician Steven Evans).⁶ Using 22 phonological, 13 morphological and 259 lexical features as coded characters, they were able to produce a tree with a “perfect phylogeny” algorithm that tracked the branching of twenty-four ancient and medieval Indo-European languages. However, the phylogeny was not quite perfect since the position of Germanic could not be determined. As it turned out from subsequent work, which included language contact (Nakleh, Ringe & Warnow 2005), this was due to the fact that Germanic was apparently in contact with the other branches and therefore did not fit the “perfect phylogeny”.

Probably the most controversial computational approach to the subgrouping and chronology of the Indo-European languages was that adopted by Russell Gray and Quentin Atkinson (2003). In their research, Gray and Atkinson used the word lists of basic vocabulary for eighty-seven Indo-European languages compiled by Dyen et al. (1992) along with Dyen et al.’s cognancy judgements and applied Bayesian inference to establish the dates for linguistic divergence of the languages analysed. They employed the algorithms for estimating the divergence time of DNA from evolutionary biology calibrated to the dates of the languages’ known split times. Using this technique, they were able to generate a tree in which the estimated dates of divergence of the particular groups of Indo-European languages were essentially in line with Colin Renfrew’s theory on the Anatolian origin of Indo-European languages (Renfrew 1987). The method was further expanded using phylogeographic approaches by Bouckaert et al. (2012), with the results also pointing to an Anatolian origin.

The work of Gray & Atkinson (2003) and Bouckaert et al. (2012) was challenged by a team from the University of California, Berkeley (Chang et al. 2015). They tried to use the same method but with the addition of ancestry constraints, i.e. information relating to the fact that Latin is the parent language of Romance etc. Their research indicated that the chronology of the Indo-European splits was significantly shorter than previously thought and roughly in accordance with the dating of the so-called steppe hypothesis in archaeology as regards the homeland of the speakers of Proto-Indo-European. Although the authors claim that “the agreement between our findings and the independent results of other lines of research confirms the reliability of statistical inference

⁶ I will outline here only the three main applications used for the classification, subgrouping and chronology of Indo-European languages as represented in Ringe et al. 2002, Gray & Atkinson 2003 and Chang et al. 2015. I refer the reader to Pereltsvaig & Lewis 2017 and McMahon & McMahon 2006 for a more in-depth analysis. A summary and assessment of most of these approaches was also given by Ringe (2017b: 67–71).

of reconstructed chronologies” (Chang et al. 2015: 194), there are problems inherent in the Bayesian inference itself (cf. Ringe in Chapter 4 and the comments by Nichols & Warnow 2008: 785 on Bayesian methods).

At present, it is difficult to say whether computational methods can give us a reliable chronology of the splits of the individual branches.⁷ It is clear, though, that they can provide us with a reliable indication of how the trees (topology) branched, as the work by Ringe et al. (2002) has shown. It is also striking that most of the approaches identify the traditionally assumed sub-groups (Indo-Iranian, Balto-Slavic, Italo-Celtic, Graeco-Armenian). Even with careful calibration, perhaps the best we can hope for is a very rough estimate. However, even this should be corroborated by linguistic, archaeological, genetic and historical data.

3.5 Computational Replication of Sound Change (Computerised Forwards Reconstruction)

Another approach which could yield promising results in the reconstruction of the Indo-European family tree, especially with regard to the relative chronology of sound changes, is the replication of sound change or computerised forwards reconstruction (cf. Sims-Williams 2018).⁸ The procedure of reconstructing forwards is not unknown in traditional historical linguistics. For example, Calvert Watkins (1962: 5) mentions it in his monograph on the Indo-European origins of the Celtic verb, where he employs the method to see how a Proto-Indo-European form would be regularly continued in Celtic. The method was also adopted by Ives Goddard (1998: 183) in his analysis of the Arapaho historical phonology, which, like the phonology of the insular Celtic languages, underwent some radical changes.

The aim of such an approach, enhanced by using a computer, is to employ an algorithm that reads the given input data (proto-forms), makes the appropriate changes based on the changes that are usually assumed to have taken place in the development of the particular languages (regular sound changes) and generates the output which can be either manually or automatically compared with the actually attested forms. The purpose of this approach is to test the regularity of sound changes and their relative chronology. If the generated form is exactly the same as the attested one, then the relative chronology is assumed to be correct.

However, only two programs known to the author have tried to replicate sound changes comprehensively by taking into account most of the changes (Maniet 1985; Hartman 2003). Most of the other programs have only used

⁷ See also Chapter 4 by Ringe on the limits of computational cladistics.

⁸ This is also sometimes called “historical derivation”; cf. Kondrak 2002: 12–15.

fragmentary data and applied only the main sound changes. I will list and discuss the most important ones below.⁹

The first application of historical derivation was a program by Raoul Smith (1969) which applied twenty-one regular rules to 650 Proto-Indo-European reconstructed lexemes as taken from an etymological dictionary and, through the application of those rules, derived the modern Russian forms from the Proto-Indo-European ones. However, only nine lexemes were derived regularly because too few sound-change rules were applied.

Further attempts were made by Burton-Hunter (1976) to generate Old French forms from their Latin sources and by Eastlack (1977), to produce Ibero-Romance from its Latin source. Both programs yielded a high percentage of expected forms since they took into account a larger number of rules within a shorter period of time (Latin to Old French or Old Spanish).

Probably one of the most comprehensive computational sound-change replication programs was the one devised by Albert Maniet (1985). He simulated the changes from Proto-Indo-European to Latin (252 rules) on a corpus of approximately 15,000 words from Plautus. However, his research went largely unnoticed by both linguists and computer scientists, and the program is, unfortunately, no longer accessible today.

In more recent times, in his thesis on cognate alignment and reconstruction, Kondrak (2002: 141–3) included a short appendix on a Perl program which generated Polish forms from their Proto-Slavic sources. From the 626 lexemes taken into account, 72.5 per cent were regular.

Hartman (2003) has been developing a program since the 1980s which simulates the sound changes from Latin to Spanish (with approximately 122 rules and 1,806 coded vocabulary items) using sets of distinctive features (from Chomsky & Halle 1968) coded as binary strings rather than the usual string substitution as in most programs developed so far. As input, the etymon is “fed” into the computer via the keyboard or from a file. Then the individual letters are translated into sets of features, and changes are applied to the features in accordance with the programmed rules and their relative chronology. Finally, the features are translated back into characters and displayed using the International Phonetic Alphabet.

Hartman also enabled the program itself to be used to model sound changes in other languages. Recently, the hard-coded version with the Spanish model was made available on-line, and a new working version of the program was presented (Hartman 2018) but with no significant modifications. The earlier version of Hartman’s program was used by Towhid bin Muzzafar to simulate the changes from Proto-Algonquian to Shawnee (bin Muzzafar 1997). Apart

⁹ I also refer the reader to the enumeration and short descriptions of the programs by Sims-Williams in his overview (2018).

from some additional examples that he was able to find for Shawnee and minor improvements in the relative chronology of changes, he pointed out an important aspect of the computational replication of sound change which probably constitutes the main obstacle to this kind of work:

It has been known that if diachronic sound change is regular, then it must be possible to demonstrate the regularity of sound change in computer models. But very few have actually ventured to take historical sound change rules from textbooks of well studied languages and develop a working computer model. And anyone who HAS ventured into this territory has quickly realized that there is a world of difference between the rules as they are written in standard linguistic notation and as they need to be written in computer models. (bin Muzzafar 1997: 73–4)

Some more recent programs, apart from the ones developed by people to create fictitious languages (so-called “Conlangers”), include the simulation of Spanish from Latin by Marcel Schmuki (“ETYMO”, 2001), Proto-Germanic from Proto-Indo-European by Brett Kessler (“Derive”, 2004), Gothic from Proto-Indo-European by Roland Mittmann (2009) and the “Sound change transducers” by Amir Zeldes (2008).

Sims-Williams tried some computational replication of sound change in Celtic historical phonology just by using the “Find & Replace” function in the word processor (Sims-Williams 2018: 562). By applying forty-three sound changes on the material of 159 selected Common Celtic forms, he was able to find amendments in the relative chronology of changes and identify usually overlooked Celtic cognates from a Proto-Indo-European root. He further argued that with modern programs, this kind of research could be pursued and expanded significantly so that it might include complete Proto-Indo-European reconstructions as input to find new etymologies, correct existing ones and check the relative chronology of sound changes (Sims-Williams 2018: 564). I agree with the author in principle, although I think implementing such an expanded approach is quite complicated, because of several key factors:

1. circularity of the method – there are competing hypotheses on what the Proto-Indo-European reconstructed forms should look like, and the applied sound changes can be used to fit the proto-form
2. problems in “translating” the sound changes into computational notation
3. problems in coding the forms and characters as the input for the program
4. lack of high-quality computational databases for Indo-European languages and Proto-Indo-European which would also include verbal and nominal paradigms
5. lack of monographs and works that comprehensively present the complete relative chronology of changes from Proto-Indo-European to the attested languages

6. not applying morphological changes which are usually assumed to have been very numerous in paradigms and which will inevitably blur the regular outcomes from simulations, especially if the simulation has a very large time span (e.g. from Proto-Indo-European to Latin etc.).

To give just one example, if we take the Proto-Indo-European reconstructed form for ‘foot’ (cf. (Wodtko et al. 2008: 530 n. 2) and apply the sound changes that occurred in Attic Greek (cf. Liesner 2015: 110–15), we will get the following result:

nom.sg.	<i>*pód-s</i>	>	<i>*pots</i>	>	<i>*pos</i> (preserved in cpd. <i>τρίπος</i> ‘tripod’)
gen.sg.	<i>*péd-s</i>	>	<i>*pets</i>	>	<i>*pes</i>

Apart from the fact that some forms do not match the attested ones because of not being included in the morphological changes, we will encounter problems with the reconstruction of the proto-form, since different scholars propose competing forms: nom.sg. **póds*, gen.sg. **pedés* (Ringe 2017a: 59) or nom.sg. **póds*, gen.sg. **pedós* (Clackson 2007: 72). There are also problems with the assumed sound changes (cf. Szemerényi 1996: 116 for the view that **póds* developed into **póss* and Greek **pós*).

3.6 Potential Ways to Enhance the Computational Replication of Sound Change

In this section, I will try to address some of the problems involved in computational replication of sound change.

3.6.1 Circularity and the Assumption of Different Reconstructed Input Forms

Probably the easiest way to avoid circularity would be to test the various competing hypotheses, which would be fairly easy if the program were interactive and allowed changes to the rules and the input forms. However, the correct form might still be different from any of those proposed so far. Therefore, what remains is to try to use an algorithm that reconstructs the proto-form or the one that would infer the changes (based on the ones that are known e.g. from Latin to Romance) rather than project them mechanically in a replicatory manner (cf. Anderson, List & Tresoldi 2018). The exact mechanism of this approach has not been fully presented yet, but if it turns out to be successful, it could prove an additional help to our understanding of linguistic changes.

3.6.2 *Problems in “Translating” the Sound Changes into Computational Notation*

This is one of the most important obstacles, since the linguistic changes cannot simply be used in computational algorithms but have to be “translated” into computational terms. That is, either as simple string substitutions as in most programs or with the use of distinctive features through a binary matrix or parsing.¹⁰ Either way, only simple sound changes such as Latin rhotacism can be easily coded in terms of string substitutions. Problems occur with changes that operate only in certain positions in the word (since the computer does not know what a syllable is, so this requires additional coding) or, even worse, changes that depend on factors outside of phonetics and phonology (e.g. the apocope of final *-i* in Latin, which occurs in verbs and adverbs but not in nouns and could be triggered by the final position of the verb in the sentence, cf. [Hock 2012](#), if correct).

3.6.3 *Problems in Coding the Forms and Characters as the Input for the Program*

Apart from translating sound changes into computational terms, the input also has to be coded in such a way as to encompass all the necessary features depending on the required changes (accent, prosody, co-articulation etc.). This, along with the mutual compatibility of the data, was recently addressed by [List \(2017\)](#), who proposed a universal format for coding etymological data.

3.6.4 *Lack of High-Quality Databases for Indo-European Languages and Proto-Indo-European*

This is also an important problem since the outcome of a computer simulation will essentially depend on the quality of the input and the programmed rules. There are hardly any reliable databases for Indo-European studies, but there are some recent projects (see [Noyer 2016](#); [Barnett & Macdonald 2018](#)), which, once completed and made available, should remedy the situation to some extent.

¹⁰ A string is understood as either a letter or a number, in this case only a letter. An example of a programmed sound change would be Latin rhotacism, where the string substitution would be programmed as follows: change every /VsV/ sequence into /VrV/ in all forms in the database. The vowel (V) would be defined as either /a(:)/, long or short respectively, /e(:)/, /i(:)/, /u(:)/, /o(:)/. The algorithm would then proceed to change every VsV sequence into VrV automatically and without exceptions, thus replicating what is usually thought of as an example of regular sound change.

3.6.5 *Lack of Monographs with the Complete Relative Chronology of Changes*

This is more of a problem for Indo-European studies in general than it is for the computational replication of sound change. There seems to be a lack of comprehensive works which present the complete (even hypothesised) relative chronology of changes. In most publications, only the chronology of main changes is given (even in e.g. [McCone 1996](#)) or the ones that are relevant to the discussion. This has been improving in recent years (cf. [Matasović 2005](#); [Olander 2015](#): 46–67), but much work in this area remains to be done.

3.6.6 *Not Applying Analogical Changes*

Whereas it is clear how to simulate regular sound changes, analogical changes are inherently irregular and occur quite unpredictably (see [Olander 2015](#): 46). If that is so, it will be problematic to code them appropriately, other than as a substitution of the whole word within the chronology of changes: e.g. gen.sg. **pód-s* → **pód-os* with this change occurring only in this specific form. This can create problems if there are two similar forms in the paradigms and analogical change occurs in only one of them. In that event, morphological annotation of the forms would be necessary to avoid such situations.

3.7 **The Potential Use of Relative Chronology of Sound Changes in Subgrouping**

In an article devoted to the position of West Germanic, Don Ringe observed that

the chronology of changes serves two purposes. On the one hand, languages are much less likely to have undergone innovations in the same order independently by chance. On the other hand, a sequence of changes should require more time to go to completion than a similar set of unrelated changes, thus ensuring that the period of linguistic unity demonstrated by the shared changes continued for a significant period of time. ([Ringe 2012](#): 33)¹¹

If that is so, then it would be possible to use the computational replication of sound change in this area as well, depending on the quality and the availability of the data as discussed below.¹²

¹¹ Cf. the similar arguments made by [Matasović 2005](#) for Balto-Slavic.

¹² For an in-depth discussion of the particular subgroups, I refer the reader to the respective chapters in this volume.

3.7.1 *Indo-Iranian*

The relative chronology of the main Indo-Iranian sound changes along with the approximate reconstruction of Proto-Indo-Iranian seem more or less established: Lubotsky (2018: 1885–6) gives a list of ten consecutive sound changes common to Indic and Iranian. Difficulties in the computational simulation include differing opinions between scholars on the place of Bartholomae's Law in the relative chronology of changes (was it also a Proto-Indo-European process or solely Indo-Iranian or perhaps two independent changes?) or the exact conditioning of Brugmann's Law and reconstruction of the proto-forms accordingly (e.g. the Proto-Indo-European reconstruction $*h_3éyis$ or $*h_2óyis$ 'sheep' depending on the assumptions made about the conditioning of Brugmann's Law and the absence of its operation in this word either due to the full grade of the ablaut or analogical change).

3.7.2 *Balto-Slavic*

There is considerable discussion about the relative chronology of Slavic and Baltic sound changes, but, although there are differences between scholars on the details and the exact relative chronology (cf. Matasović 2005; Kortlandt 2008; Olander 2015), the main sound changes seem to be established, indicating the existence of a subgroup: Olander (2015: 47–53) lists eleven consecutive sound changes common to Balto-Slavic. Technical problems arise in the computational simulation with the "translation" of the changes in computerised terms concerning the Balto-Slavic accentuation and coding of the accent. Furthermore, there is a problem with the double reflex of Proto-Indo-European syllabic resonants in Balto-Slavic as either $*iR$ or $*uR$, since the exact conditioning has to be stated for the computational simulation.

3.7.3 *Italo-Celtic*

For Italo-Celtic, sound changes seem relatively less important than morphological changes (cf. de Vaan 2008: 7; Weiss 2020: 493–5). Weiss (2020: 207) gives a list of four consecutive sound changes that could be common to Italic and Celtic. There are also controversies concerning whether this stage existed at all – compare the arguments of Meiser (2003: 30–1) and the discussions by Schrijver (2006: 48–53) and also recently by Zair (2018). There is hardly any complete hypothesis on the reconstruction of Proto-Italo-Celtic (cf. Kortlandt 2007: 149–78) or even a balanced account of Proto-Italic (cf. van der Staaïj 1995).

3.7.4 *Graeco-Armenian*

The relative chronology of changes usually postulated for Ancient Greek and Armenian does not seem to support a Graeco-Armenian subgroup (cf. [Kim 2018](#)). Mostly lexical items favour this grouping. It is also the weakest in the computational cladistic analysis by [Ringe et al. 2002](#).

From the point of view of the chronology of sound changes, only the Indo-Iranian and Balto-Slavic subgroups appear to be real entities. Proto-Indo-Iranian and Proto-Balto-Slavic also have more or less established reconstructions. However, in order to be able to fully investigate the relative chronology of sound changes, it would be necessary to compile a comprehensive list, have the changes translated as closely as possible from linguistic to computational notation and use a high-quality database with more or less complete data.

Since it has long been a gold standard in historical linguistics that morphological innovations should be taken more seriously into account than phonological or lexical ones, in the [next section](#), I will discuss the possibilities and perspectives of including morphological changes in the computational replication of sound change.

3.8 Perspectives on the Inclusion of Morphological Changes

It may be possible to expand the scope of the computational replication of sound change in such a way as to apply computationally generated sound changes along with morphological changes to the complete lexicon of the Proto-Indo-European language as it is reconstructed today in order to generate the main Indo-European languages. The program would apply the sound changes and the analogical changes in their relative chronology to the lexicon of Proto-Indo-European and generate output which in turn would be compared with the actual data relating to those languages. With this approach, the amount of regular sound change from a more or less complete lexicon would be uncovered along with the exact interferences causing irregularities – errors in the formulation, chronology or translation into computational terms of the programmed sound changes, borrowing and, especially, analogy.

This approach could potentially address a very direct and practical question of interest to every practising historical linguist: whether one analogical solution is more probable than another. The usual answer to this question depends on one's view of the system of the language in which the change occurred. However, different scholars might view a certain language system (in its earlier or even reconstructed phases) differently and so pose different analogical explanations with their own models and motives. They will look for parallel developments and typologically similar changes in the material they are working on and, most importantly, in their previous experiences. We can deem one

solution of analogical remodelling as more plausible than another by providing other analogical solutions of exactly the same type, with similar models and motives along with an in-depth analysis of the synchronic situation. Warren Cowgill, in his work on universals in Indo-European diachronic morphology, noted that with regard to small-scale innovations “[a] sufficiently large collection of such individual changes, appropriately classified, should give linguists measure of the relative plausibility of different solutions for problems in historical grammar” (1966: 115). He continues:

At present each linguist judges the plausibility of a newly proposed solution pretty much by what he happens to remember of the morphologic innovations which during his career he has been led, for one reason or another, to accept as plausible. A reasonably objective standard of plausibility should make it easier for historical linguists to agree on solutions for problems of historical morphology that at present are still disputed. (Cowgill 1966: 115–116)

Using a computational algorithm and an electronic database with word forms and the phonological and morphological changes which occurred in their development, it would be possible to create a virtually complete picture of the phonological and morphological system of the language at every stage of its development and to investigate any possible phonetic and analogical changes. Because the reconstruction of the proto-language and each of its stages of development remains hypothetical, its validity and accuracy can only be checked against the general typology of both synchronic language systems and types of diachronic changes along with the internal coherence of the system. Most notably, the compatibility of every single sound change can be checked against the hypothesis by applying it to the lexicon of the language.

Such an approach would allow us to formulate hypotheses concerning the relative chronology and tendencies of sound change and analogical levelling based on fairly complete empirical data. The results would confirm or challenge the existing theories on sound change and analogical remodelling and could form the basis for comprehensive historical grammars in the future which, with the expansion of integrated corpus linguistics, could encompass all corpora of texts from all periods of the documented language development. Such a large database would enable scholars to pursue further research in the area, allow the explicit discussion of competing hypotheses and serve as an educational tool. Additionally, the method itself could be applied to other language families, thus forming the basis for research on universal tendencies in language change. Moreover, it would break the so-called “handbook tradition” mentioned by Eichner (1992: 61), whereby a sound change is illustrated only by a handful of examples (usually the same in various historical grammars) and in order to find more of these, one has to consult an etymological dictionary.

In fact, a recent project carried out by Jouna Pyysalo (2017) has managed to achieve most of what is described above. With the use of computational simulation, the project aims to generate all the forms of the Indo-European languages from their reconstructed proto-forms. However, the author uses an idiosyncratic reconstruction of Proto-Indo-European which deviates significantly from the current *communis opinio*, i.e. with only one laryngeal, at the same time basing his argument against the classical laryngeal theory only on Anatolian data and completely ignoring the data relating to Ancient Greek (cf. Janhunen & Pyysalo 2018). If the input proto-forms are hard-coded so that they cannot be modified, this project will only serve to present the author's own views on the subject.

3.9 Perspectives on the Computational Methods

Czekanowski, Kroeber and Chrétien pursued an interesting way of handling language classification back in the days when such methods were far from popular or even acceptable. They were often very harshly criticised by linguists, to the extent that virtually nobody followed their lead to improve the method. Just as Czekanowski, Kroeber and Chrétien were pioneers in the use of statistics for language classification applied to Indo-European, so were Ringe et al. and Gray and Atkinson pioneers in the use of computational methods in the same area. Even though their work is relatively recent, a large amount of new research has been done in the field, which has become so popular that some scholars argue that historical linguistics appears to have taken something of a quantitative turn. Indeed, new methods are being implemented in an attempt to meet the standards of traditional historical linguistics in paying careful attention to the annotation of the data and to detail in general, while at the same time taking advantage of the replicability, robustness and formality of the computational approach. Progress is being made in modelling language characteristics and change using algorithms, and more attention is being given to making the programs and the data openly accessible, in a more or less standard format, easy to use by non-computational scientists and, importantly, annotated jointly by experts in the respective fields. It seems that only through combining qualitative and quantitative methods can further progress be made in the field of Indo-European linguistics, and the current thinking is that such progress is only possible if scholars from different disciplines contribute collectively.

The advantages of the computational approach are its speed, error-free processing of data and ability to handle large amounts of data. I would further argue that, thanks to the computational approach and the explicit presentation of the material, it would be much easier to compare different linguists' competing hypotheses, even for people from outside the exact field of

specialisation, thus making the whole enterprise of Indo-European linguistics easily accessible for interdisciplinary studies.

References

- Anderson, Cormac, Johann-Mattis List & Tiago Tresoldi. 2018. Modelling sound change with the help of multi-tiered sequence representations. Paper presented at the 48th Poznań Linguistic Meeting, Poznań, 15 September 2018. <https://bit.ly/3NtRK7F> (accessed 10 January 2020).
- Baker, Adam. 2008. Computational approaches to the study of language change. *Language and Linguistics Compass* 2. 289–307.
- Barnett, Phillip & Ryan Macdonald. 2018. DERBiPIE: A computational database for Proto-Indo-European language. Paper presented at the Thirtieth Annual UCLA Indo-European Conference, November 9–10, 2018, University of California, Los Angeles.
- Bergsland, Knut & Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3. 115–53.
- Bouckaert, Remco, et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337. 957–60.
- Burton-Hunter, Sarah K. 1976. Romance etymology: A computerized model. *Computers and the Humanities* 10. 217–20.
- Campbell, Lyle. 2013. *Historical linguistics: An introduction*. 3rd ed. Edinburgh: Edinburgh University Press.
- Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91. 194–244.
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Chrétien, Charles Douglas & Alfred Kroeber. 1937. Quantitative classification of Indo-European languages. *Language* 13. 83–103.
- Clackson, James. 2007. *Indo-European linguistics: An introduction*. Cambridge: Cambridge University Press.
- Cowgill, Warren. 1966. A search for universals in Indo-European diachronic morphology. In Joseph Greenberg (ed.), *Universals of language*. 2nd ed., 114–41. Cambridge, MA: MIT Press.
- Czekanowski, Jan. 1927. *Wstęp do historii Słowian [Introduction to the history of the Slavs]*. Lwów: Jakubowski.
- Dunn, Michael. 2015. Language phylogenies. In Claire Bowern & Bethwyn Evans (eds.), *Routledge handbook of historical linguistics*, 190–211. Abingdon: Routledge.
- Dyen, Isidore, Joseph Kruskal & Paul Black. 1992. An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82. 1–132.
- Eastlack, Charles L. 1977. Iberochange: A program to simulate systematic sound change in Ibero-Romance. *Computers and the Humanities*, 11. 81–8.
- Eichner, Heiner. 1992. Indogermanisches Phonemsystem und lateinische Lautgeschichte. In Oswald Panagl & Thomas Krisch (eds.), *Latein und Indogermanisch*, 55–79. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Embleton, Sheila. 1986. *Statistics in historical linguistics*. Bochum: Brockmeyer.

- Goddard, Ives. 1998. Recovering Arapaho etymologies by reconstructing forwards. In H. Craig Melchert & Jay H. Jasanoff (eds.), *Mir curad: Studies in honor of Calvert Watkins*, 183–200. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Gray, Russell & Quentin Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–9.
- Greenhill, Simon. 2011. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37. 689–98.
- Hartman, Steven Lee. 2003. Phono (Version 4.0): Software for modeling regular historical sound change. In Leonel Ruiz Miyares, Celia E. Álvarez Moreno & María Rosa Álvarez Silva (eds.), *Actas: VIII Simposio Internacional de Comunicación Social*. Vol. 1, 606–9. Málaga: Universidad de Málaga & Santiago de Cuba: Centro de Lingüística Aplicada.
- Hartman, Steven Lee. 2018. Phono, software for historical phonology. <https://langnhist.weebly.com/phonoTOC.html> (accessed 10 January 2020).
- Hewson, John. 1974. Comparative reconstruction on the computer. In John M. Anderson & Charles Jones (eds.), *Historical linguistics*. Vol. 1. *Syntax, morphology, internal and comparative reconstruction*, 191–7. Amsterdam: North-Holland.
- Hock, Hans. 2012. Phrasal prosody and the Indo-European verb. In H. Craig Melchert (ed.), *The Indo-European verb*, 115–26. Wiesbaden: Reichert.
- Hoijer, Harry. 1954. Lexicostatistics: A critique. *Language* 32. 49–60.
- Jäger, Gerhard. 2018. Computational historical linguistics. ArXiv:1805.08099. <https://arxiv.org/abs/1805.08099> (accessed 10 January 2020).
- Janhunen, Juha & Jouna Pyysalo. 2018. Indo-European linguistics in the 21st century (1): From trilaryngealism to monolaryngealism – returning to Oswald Szemerényi. *Urindogermanische Sprachforschungen* 2018(1). 1–14.
- Kessler, Brett. 2001. *The significance of word lists*. Stanford: CSLI.
- Kessler, Brett. 2004. Derive. <http://spell.psychology.wustl.edu/derive/> (accessed 10 January 2020).
- Kim, Ronald I. 2018. Greco-Armenian: The persistence of a myth. *Indogermanische Forschungen* 123. 247–72.
- Kondrak, Greg. 2002. *Algorithms for language reconstruction*. Unpublished PhD thesis, University of Toronto.
- Kortlandt, Frederik. 2007. *Italo-Celtic origins and prehistoric development of the Irish language*. Amsterdam: Rodopi.
- Kortlandt, Frederik. 2008. Balto-Slavic phonological developments. *Baltistica* 43. 5–15.
- Liesner, Malte. 2015. *Greek historical phonology workbook*. Wiesbaden: Reichert.
- List, Johann-Mattis. 2017. A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets. In *Proceedings of the EACL 2017 Software Demonstrations*, 9–12. Stroudsburg, PA: Association for Computational Linguistics.
- Lubotsky, Alexander M. 2018. The phonology of Proto-Indo-Iranian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 1875–88. Berlin: De Gruyter Mouton.
- McCone, Kim. 1996. *Towards a relative chronology of ancient and medieval Celtic sound change*. Maynooth: Department of Old Irish, St. Patrick's College.

- McMahon, April & Robert McMahon. 2006. *Language classification by numbers*. Oxford: Oxford University Press.
- Maniet, Albert. 1985. Un programme de phonologie diachronique: De l'“indoeuropéen” au latin par ordinateur – version définitive. *Cahiers de l'Institut de Linguistique de Louvain* 11. 203–43.
- Matasović, Ranko. 2005. Toward a relative chronology of the earliest Baltic and Slavic sound changes. *Baltistica* 40. 147–57.
- Meillet, Antoine. 1922. *Les dialectes indo-européens*. 2nd ed. Paris: Champion.
- Meiser, Gerhard. 2003. *Veni, vidi, vici: Die Vorgeschichte der lateinischen Perfektsystems*. Munich: Beck.
- Mittmann, Roland. 2009. Ein Verfahren zur Ermittlung der relativen Chronologie der vorgotischen Lautgesetze. In Christian Chiarcos, Richard Eckart de Castillo & Manfred Stede (eds.), *Von der Form zur Bedeutung: Texte automatisch verarbeiten / From form to meaning: Processing texts automatically*, 199–209. Tübingen: Narr.
- bin Muzzafar, Towhid. 1997. *Computer simulation of Shawnee historical phonology*. Unpublished MA thesis, Memorial University of Newfoundland.
- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81. 382–420.
- Nichols, Johanna & Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2. 760–820.
- Noyer, Rolf. 2016. *Etymological database for Indo-European languages*. <https://bit.ly/3wJU5EK> (accessed 10 January 2020).
- Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. Leiden: Brill.
- Pereltsvaig, Asya & Martin W. Lewis. 2017. *The Indo-European controversy: Facts and fallacies*. Cambridge: Cambridge University Press.
- Pyyssalo, Jouna. 2017. Proto-Indo-European lexicon: The generative etymological dictionary of Indo-European languages. In Jörg Tiedemann & Nina Tahmasebi (eds.), *Proceedings of the 21st Nordic Conference of Computational Linguistics*, 259–62. Linköping: Linköping University Electronic Press.
- Renfrew, Colin. 1987. *Archaeology and language: The puzzle of Indo-European origins*. London: Penguin.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Ringe, Don. 2012. Cladistic principles and linguistic reality: The case of West Germanic. In Philomen Probert & Andreas Willi (eds.), *Laws and rules in Indo-European*, 33–42. Oxford: Oxford University Press.
- Ringe, Don. 2017a. *From Proto-Indo-European to Proto-Germanic*. 2nd ed. Oxford: Oxford University Press.
- Ringe, Don. 2017b. Indo-European dialectology. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 1, 62–75. Berlin: De Gruyter Mouton.
- Safarewicz, Jan. 1948. Krytyka metody ilościowej stosowanej w ocenie pokrewieństwa językowego [A critique of the quantitative method used in evaluating linguistic relationship]. *Biuletyn Polskiego Towarzystwa Językoznawczego* 8. 30–9.
- Schmuki, Marcel. 2001. ETYMO. <http://www.etymo.net/> (accessed 10 January 2020).

- Schrijver, Peter. 2006. Review of Meiser 2003. *Kratylos* 51. 46–64.
- Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.
- Sims-Williams P. 2018. Mechanising historical phonology. *Transactions of the Philological Society* 116. 555–73.
- Smith, Raoul N. 1969. A computer simulation of phonological change. *ITL: Tijdschrift voor Toegepaste Linguïstiek* 5. 82–91.
- van der Staaij, Robert. 1995. *A reconstruction of Proto-Italic*. Unpublished PhD thesis, Leiden University.
- Swadesh, Morris. 1952. Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96. 452–63.
- Szemerényi, Oswald. 1996. *Introduction to Indo-European linguistics*. Oxford: Oxford University Press.
- Tischler, Johann. 1973. *Glottochronologie und Lexikostatistik*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- de Vaan, Michiel. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.
- Watkins, Calvert. 1962. *Indo-European origins of the Celtic verb*. Dublin: Dublin Institute for Advanced Studies.
- Weiss, Michael. 2020. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor, MI: Beech Stave.
- Wichmann, Søren, Eric Holman & Cecil Brown (eds.). 2018. The ASJP database (version 18). <http://asjp.clld.org/>.
- Wodtko, Dagmar, Britta Irslinger & Carolin Schneider (eds.). 2008. *Nomina im indo-germanischen Lexikon*. Heidelberg: Winter.
- Zair, Nicholas. 2018. The shared features of Italic and Celtic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.). 2018. *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 2030–7. Berlin: De Gruyter Mouton.
- Zeldes, Amir. 2008–9. Sound change transducers. <https://corpling.uis.georgetown.edu/amir/sct.php> (accessed 10 January 2020 but no longer available).

4 What We Can (and Can't) Learn from Computational Cladistics

Don Ringe

For more than twenty years now various teams of colleagues have been pursuing computational work on the cladistics of Indo-European. I am partly to blame, since my collaboration with Tandy Warnow helped to make such research visible and attractive. To at least some observers, it has not always been clear that what we can learn from computational cladistics is limited. This chapter is an attempt to explore those limits.

4.1 Outgroup Analysis

I begin with a well-known principle of traditional cladistics that should be kept in mind as background for a consideration of computational methods, namely outgroup analysis. A simple example is given in [Figure 4.1](#).

The reflexes of PGmc. **tūna-* ‘enclosure’ are always *a*-stems, reflecting pre-Proto-Germanic *o*-stems, but they are neuter in Norse and masculine in West Germanic, so the gender of the proto-form cannot be recovered by evidence internal to Germanic. The reflexes of the corresponding word in Celtic are always neuter, but the Old Irish word is an *s*-stem, while the Gaulish word is an *o*-stem – at least to judge from the Latinized form recorded in place names. Leaving that last problem aside (since this is just a demonstration of method), we would have to say that the gender, but not the stem class, of the Proto-Celtic form can be recovered by internal evidence. But if the two problems are considered together, the simplest solution is that the earliest recoverable form of the word was **dūnom*, a neuter *o*-stem, because that hypothesis requires only two changes: a shift of gender in West Germanic, and a shift of stem class in Old Irish.¹

I am grateful to Bob Berwick and Tandy Warnow for helpful discussion of various parts of this chapter. All errors and infelicities are mine.

¹ Note that this conclusion is valid regardless of whether the Celtic and Germanic words reflect common inheritance or early borrowing of a Celtic word into Proto-Germanic. I emphasize that because it is one illustration of an important point: subgrouping and establishing a genetic relationship (of words or languages) in the first place are different problems, and they cannot be solved by the same methods.

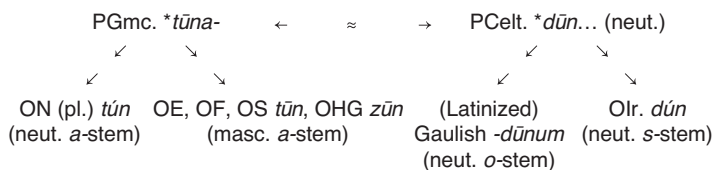


Figure 4.1 Outgroup analysis of PGmc. *tūna- and PCelt. *dūn ...

4.2 Computational Cladistics

In this textbook illustration, we took the shape of the cladistic tree for granted as a basis for investigating another type of problem. Cladistics inverts that, using details of the linguistic data to find the true tree.² Nevertheless, to a large extent (though not completely), a problem in computational cladistics uses the same mathematical principle as outgroup analysis. The most widely employed criterion for tree optimization – that is, for choosing the best of the trees that the software returns – is maximum parsimony: the optimal tree is the tree on which the smallest number of individual changes is required to account for the observed data. That is essentially the line of reasoning employed in the illustration above. Alternative criteria can be (and are) employed. For instance, the maximum compatibility criterion looks for the tree on which the greatest number of characters (that is, words or features) are “compatible” with the tree, i.e. the maximum number which fit the tree with no parallel development and no backmutation. In principle the two criteria are quite different. The maximum compatibility criterion can yield an optimal tree in which there is a great deal of backmutation and parallel development so long as it’s confined to, say, 1 per cent of the words in the comparative wordlist; they can be as messy as you like, as long as there are only a few of them. Maximum parsimony yields the tree with the smallest amount of mess overall, regardless of how it’s distributed. But in practice the two criteria usually give similar results, and if the amount of parallel development and backmutation in a dataset is very small, the results of the two methods converge.

Cladistics involves more than the inverse of outgroup analysis, however. For one thing, automation is necessary because of the sheer size of the problem. As the number of languages compared – in cladistic terms, the number of taxa – increases, the number of possible binary-branching trees that must be considered increases exponentially. If n is the number of taxa, the numbers of possible rooted and unrooted binary branching trees are given by the formulae in Table 4.1 (Dobson [no date]; Embleton 1986: 28–9 with references).

² For an introduction to computational cladistics and the terminology that computational cladists use, see Nichols & Warnow 2008.

Table 4.1 *Number of rooted and unrooted binary branching trees (n = number of taxa)*

Number of distinct rooted binary trees:	$(2n - 3)(2n - 5) \cdots 5 \cdot 3 \cdot 1$
Number of distinct unrooted binary trees:	$(2n - 5) \cdots 5 \cdot 3 \cdot 1$

If the problem under investigation is large enough to be interesting, it’s not just that one human lifetime is too short to do the calculations by hand (though that can be true); it’s also that the human mind can’t keep track of all the possibilities. The computer, of course, can.

4.3 Problems with Character Evidence

Traditional cladistics is also beset by another problem that computers are ideally suited to solve. Consider the types of characters used in linguistic cladistics. Lexical characters (vocabulary) are actually the least reliable, because parallel semantic development is rampant – words meaning ‘person’ often come to mean ‘man’ and then ‘husband’, for instance – and undetectable borrowing between closely related languages is a real problem. Moreover, we expect phonological and morphological characters to give a better picture of linguistic descent because they are grammatical, and grammar is acquired in native language acquisition in the first few years of life and resists modification later in life. But phonological and morphological characters have weaknesses of their own as well as strengths. Even if they are based on mergers (not simply on phonetic changes), phonological characters are usually “natural” and easily repeatable, making parallel development a significant problem (Ringe, Warnow & Taylor 2002: 66–7; Ringe & Eska 2013: 257–9); their strength is that mergers are irreversible, which means that the direction of a tree edge in time can be established. By contrast, changes in inflectional morphology are hardly ever repeatable in detail (except for loss of a morphological category or marker, which occurs often); but it is often difficult to figure out which state of a morphological character is original and which states are innovative.

Of course there are traditional ways around these problems. Though the probability of any single sound change recurring independently is usually fairly high, the probability of a whole set of sound changes – especially an ordered set – recurring independently is far lower. The most distinctive sound changes that define the Germanic subgroup are a case in point. The following seven interrelated sound changes occurred in the prehistory of every well-attested Germanic language (Ringe 2017: 113–27, 147–50):

- a. PIE $*p t k k^w >$ fricatives $*f \theta x x^w$ unless an obstruent immediately preceded;
- b. PIE $*b d g g^w > *p t k k^w$ simultaneously with or after (a);
- c. PIE breathy-voiced $*b^h d^h g^h g^w >$ fricatives $*\beta \delta \gamma \gamma^w$;
- d. $*f \theta s x x^w > *b \delta z \gamma \gamma^w$ if not word-initial *and* not adjacent to a voiceless sound *and* the last preceding syllable nucleus was unaccented ("Verner's Law"); must have followed (a), which fed it;
- e. $*\beta \delta \gamma \gamma^w > *b d g g^w$ after homorganic nasals, and $*\delta > *d$ also after $*l$ and $*z$ (at least); must have followed both (c) and (d), which fed it, and also (b), which it counterfed; further, $*\beta \delta > *b d$ word-initially, which likewise must have followed (b) for the same reason;
- f. stress was shifted to the first syllable of the word; this must have followed (d), because it both created and destroyed triggering environments for (d);
- g. unstressed $*e > *i$ unless $*r$ followed immediately; must have followed (f), which both fed and bled it.

We have no basis for *calculating* the probability that each sound change would occur in a given line of descent within a given time period, but it turns out that that does not matter, because a Bayesian approach to probabilities will yield an overall result in the right ballpark. Let us *estimate* the probability of each sound change, do the relevant calculation, and try to assess the results (see [Ringe & Eska 2013](#): 259–61).

- (a) or something very like it, occurred also in Armenian, thus in two of the ten well-attested subgroups of IE; let us therefore assign it a probability of 0.2 (two in ten);
- (b) occurred also in Armenian and Tocharian, so we assign it a probability of 0.3;
- (c) might have occurred also in Proto-Italic ([Meiser 1986](#): 38), so we assign it a probability of 0.2;
- (d) or a very similar change, occurred in fifteenth-century English ([Jespersen 1909](#): 199–208); given its complexity, we might assign it a probability of 0.1;
- (e) is commonplace (cf. the allophones of voiced obstruents in modern Spanish) and so should be assigned a high probability, say 0.5;
- (f) also occurred in Proto-Italic and Proto-Celtic, so we assign it a probability of 0.3;
- (g) is a common and repeatable merger, so a probability of 0.5 is again reasonable.

Using these crude estimates, we can calculate the probability that all seven of these sound changes would occur in a single line of descent *by chance* as

$$0.2 \times 0.3 \times 0.2 \times 0.1 \times 0.5 \times 0.3 \times 0.5 = 0.00009, \text{ or about one in } 11,111.$$

Of course, our estimates of the individual probabilities might be inaccurate. But *because they are all between 0.1 and 1*, the estimated cumulative probability

cannot be more than about an order of magnitude too small; it could easily be too large, in which case we are constructing an argument *a fortiori*. However, we are not finished with our calculation. We can establish several relative chronologies among these seven changes:

(a) $\rightarrow$ (d) $\rightarrow$ (f) $\rightarrow$ (g)

(a) $\rightarrow$ (b) $\rightarrow$ (e)

(c) $\rightarrow$ (e)

(d) $\rightarrow$ (e)

Consider only the first and longest of those chronologies. The sound changes involved could have occurred in any order, yet they did occur in this one. The number of orders in which four events could occur is $4 \times 3 \times 2 \times 1 = 24$. To account for the fact that the changes occurred in only one of the possible orders we need to divide our above result by 24, yielding 0.00000375, or about one in 266,667. Since only about 7,000 human languages are attested, the fact that all these sound changes occurred, in the chronological order reconstructible, in the prehistory of every Germanic language can only mean that they occurred once, in the common ancestor of those languages. This is an overwhelming validation of the Germanic subgroup by sound change alone.

To validate the Germanic clade, then, we do not need computational methods. Unfortunately not every potential clade offers us such clear phonological evidence; in effect, we got lucky with Germanic. Using characters based on inflectional morphology requires an even greater degree of luck: we need to find a shared morphological character state which, because of its details, is overwhelmingly likely to be an innovation. Once again Germanic is a case in point. The “weak” preterite bears a superficial resemblance to (1) the Gaulish *t*-preterite; (2) the Oscan *-tt*-perfect; (3) the Lithuanian imperfect in *-davo-*. But the details of all four formations are so different that they must have arisen independently. It follows that the weak preterite must be a Germanic innovation, and that too validates the clade. Some clades provide morphological evidence of that quality; unfortunately, many others do not.

However, computational cladistics can extract the greatest amount of information from phonological and morphological characters by combining them. We use both sets to find the best *unrooted* tree; because the tree is unrooted at this initial stage of the investigation, the fact that we might not be sure which states of morphological characters are innovative is not a problem. *Then* we use the probative phonological characters, which are usually few, to root the tree, relying on the fact that mergers are irreversible.

In principle, then, computational cladistics should be able to solve any subgrouping problem for which there is enough clear evidence in the data. Unfortunately that condition frequently remains unmet. Still worse, many datasets present the researcher with conflicting evidence. There are at least two rather different reasons for that, conceptually distinct even though they shade into one another in practice.

4.4 Phenomena Incompatible with Cladistic Trees

On the one hand, it is possible that the diversification of a family of languages simply hasn't been treelike. In that case an appropriate algorithm will find several possible trees, but none of them will be very good by any optimization criterion, and each will be bad in a different way. Early in the line of work that resulted in Ringe, Warnow & Taylor 2002, we decided to find out what such a case would look like in detail. To that end we did a cladistic analysis of some modern West Germanic languages, with Danish and Swedish as an outgroup, using PAUP*, a program designed to find the most parsimonious tree (see above). We actually expected the analysis to fail, because it's clear that most West Germanic languages have been in contact, trading material and influencing one another, for as long as they've existed; in fact ocular inspection of the data shows that there are so many overlapping patterns of cognation that no perfect phylogeny (PP, i.e. a tree in which no character exhibits parallel development or backmutation) can exist for this dataset. The computational analysis did fail spectacularly (and not only in ways that we had foreseen, because we hadn't paid enough attention to Scandinavian influence on English and Danish influence on North Frisian). Our results are given in Table 4.2.

The best possible parsimony score is simply the number of state-to-state transitions within characters; if a PP had existed, that would have been its parsimony score. The best trees that we were able to find all exhibit more than

Table 4.2 *Best possible parsimony scores for West Germanic*
Best possible parsimony score for the data: 262

Actual scores	Tree assigned each score
309	(Eng (WFrís NFrís)) (Neth HG)
313	((WFrís NFrís) (Neth HG)) Eng
315	((Eng (WFrís NFrís)) Neth) HG
319	((NFrís HG) (WFrís Neth)) Eng
329	(NFrís HG) (Eng (WFrís Neth))
335	((((WFrís HG) Neth) Eng) NFrís

forty additional state transitions, reflecting either parallel development or backmutation. For technical reasons we cannot guarantee that the algorithm found the best available tree, so in principle we cannot exclude the possibility that a closer approximation to a PP for this dataset can be found, but in practice that is highly unlikely. It can be seen that the three least bad trees are plausible: to put it in terms that are in part anachronistic, the first groups Anglo-Frisian against Franconian, the second groups English against continental West Germanic, and the third groups Ingvaëonic against High German. But their parsimony scores are all mediocre, and numerous characters are incompatible with each tree. Still worse, the next three trees have only modestly less acceptable scores but are all implausible, since all three split the Frisian languages. This is what total failure, because the diversification of a family was not treelike, looks like.³

The other possibility is that there *is* a treelike signal in the data, but that it has been obscured by undetectable borrowing between the languages. There is probably more than one way to approach that problem, but the most straightforward is to take several of the best trees and see how many “contact edges” you need to add to make *all* the data compatible with the tree. Since each contact edge must represent a historical episode of language contact, they must be posited so as to be compatible with what is known about the geography of the languages in question and the relative chronology of the family’s diversification events. Nakhleh, Ringe & Warnow 2005 is the only attempt to do that that I am aware of; interested readers should consult that work for further discussion.

Tree-networks like these can arise in more than one way in the real world, of course. “Clean speciation” followed by renewed contact and linguistic borrowing that cannot be detected (because no crucial sound changes were involved) is one way. Another possibility is that the diversification of the family was actually network-like, but only non-adjacent members of the dialect network survive; in that case the lateral edges can represent innovations which spread through the dialect network as it was diversifying, and their sparseness is simply an artefact of the originally non-adjacent positions of the survivors. In general, cladistics cannot differentiate between those two scenarios.

4.5 Time Depth in Linguistic Cladistics

Thus far I have been discussing cladistics *sensu stricto*, i.e. the recovery of the branching tree that correctly reflects a language family’s diversification. Numerous researchers have claimed that it is also possible to recover the approximate time in prehistory when each instance of diversification in a tree

³ For a good exploration of the ways in which language families can diversify, see Ross 1997, with exemplification in Ross 1998.

occurred. The most recent such claim was made by Russell Gray and his co-workers (first in [Gray & Atkinson 2003](#)) – and demolished by Andrew Garrett's team at Berkeley ([Chang et al. 2015](#)). The easiest way to discuss the problems involved in dating linguistic divergences is to discuss Gray's work.

Gray claimed that new and more powerful Bayesian cladistic methods yielded greatly improved trees and – more importantly – allowed researchers to recover the time depths of particular “speciation events” in the prehistory of language families with greater precision. He applied his methods to the Indo-European family (at first to bad data, but increasingly to competently vetted wordlists) and derived dates for PIE that are compatible with Colin Renfrew's “out of Anatolia” scenario ([Renfrew 1987](#)), but not with the “steppe hypothesis” ([Anthony 2007](#), [Anthony & Ringe 2014](#)) that most Indo-Europeanists have long believed to be most probable. Both Indo-Europeanists and computer scientists were inclined to dismiss Gray's work from the start. For one thing, Bayesian cladistics is not in any way mathematically superior to methods already available; it is merely fashionable. For another, it is not inaccurate to say that Gray took already available data and cranked them through prefabricated software. But no one would have cared about that if the work had been cogent. Unfortunately, there were always multiple reasons to suspect that it couldn't be cogent, as follows (see also [Pereltsvaig & Lewis 2015](#) for further extensive discussion).

First, Gray used only lexical data, which are the least reliable for cladistics ([Ringe, Warnow & Taylor 2002](#): 65 with references; [Nakleh et al. 2005](#)).

Secondly, *there is no lexical “clock”* – that is, the replacement of vocabulary items does not proceed at an even approximately constant rate ([Bergsland & Vogt 1962](#)). Moreover, none of the other simplifying assumptions about the rate of word replacement holds up empirically. For instance, the “rates across sites” assumption sometimes encountered in biological cladistics – namely, that if one character evolves, say, half again as quickly in lineage A as in lineage B, you can count on other characters to do the same – clearly does not hold in language development. Gray (and others who have worked in linguistic cladistics, for instance the late Isidore Dyen; see [Dyen, Kruskal & Black 1992](#)) have suggested that that need not matter: if you let the assumed rate of change vary randomly around a mean, the result will be realistic. But it's not clear that even that is loose enough; and of course the wider the variation in rates of change, the more uncertain the hypothetical dates of proto-languages become.

Thirdly, there are serious evidential problems which have an impact on the mathematics of trying to work backwards into prehistory. Steve Evans and co-authors laid out the problem in formal terms in their article of [2006](#), but it can also be stated informally (Bob Berwick, p.c.). To paraphrase Berwick, we want a theory that can infer backwards in time from a currently observed state so as to recover the dynamic processes that led to that state. In order to describe what

happened accurately, we need to know (a) the nature of the forces that have operated, (b) the magnitude of those forces, (c) the length of time over which they have operated, and (d) the initial state. In this case we are trying to derive (c), so we need to have all the other variables fixed. We linguists believe that we understand (a) well enough; but (d) is invariably full of gaps – there are some things about the proto-language that we simply cannot reconstruct because not enough evidence survives anywhere – and empirical observation shows that (b) varies within limits which are incompletely known but clearly wide. At least one further problem is the loss of data which can never be recovered, as follows. If a word *x* in a given meaning can be reconstructed securely for the proto-language, and if in the earliest records of a daughter it has been replaced by *y*, we know that *at least* one episode of replacement occurred in the unobservable prehistory of that daughter; we do *not* know whether one or more than one occurred. Thus even if the rate of vocabulary replacement were more nearly constant, we could not use it to extrapolate into prehistory with any confidence.

Fourthly, incorrect assumptions about the descent of particular languages in the tree can lead to unforeseen problems in calculating time depths. That was shown brilliantly by [Chang et al. 2015](#). They noted that, while both Latin and various Romance languages were in the database of Gray's project, the algorithm was not informed that Latin was the ancestor of the Romance languages – and likewise with Sanskrit and modern Indic, and a few other, less substantial cases. The program thus returned a tree in which Latin was the *sister* of the Romance group, Sanskrit was the sister of modern Indic, and so on. The result was to lengthen the time depths calculated from the wordlists. Chang et al. introduced constraints *forcing* the program to treat Latin as the ancestor of Romance, etc. – and the time depths shortened dramatically, yielding a date for PIE compatible with the steppe hypothesis of Anthony and others and *not* with Renfrew's "out of Anatolia" hypothesis. Gray has protested that Classical Latin is not *exactly* the ancestor of Romance, but Chang et al. replied (correctly) that if all you're using is basic wordlists, the right question is whether the Latin wordlist is the ancestor of the Romance wordlists (so far as we can tell), and the answer is clearly yes (see the extensive discussion of [Chang et al. 2015](#): 205–8). Of course this all illustrates the fact that if you want to pursue linguistic cladistics you need to have *both* a world-class linguist *and* a competent computer scientist on the team, but it also illustrates something else: the results of Bayesian cladistics are not robust; you can tweak one detail and get dramatically different results.

Finally, there is a further problem with Bayesian analyses, which was pointed out in a devastating paper by Bob Berwick ([Berwick 2015](#), unfortunately still unpublished). Berwick noticed that the "higher" nodes in Gray's best tree had low bootstrap values, often no better than 20–30 per cent. Of

course, the alternatives all had even lower bootstrap values, so the tree presented could be called the “most probable” consensus tree; but a 30 per cent probability is just not probable enough – bootstrap values that low are unacceptable to a real computational cladist. Berwick ran appropriate software on Gray's data thousands of times and superimposed all the trees returned to give a visual impression of the problem; the top of the tree was a blur, with no resolution – and that remained true even when a million iterations were run. But if you can't be sure you have the right tree, it's not feasible to estimate divergence times. Unfortunately that applies to Garrett's results no less than to Gray's, since Garrett's team set out to replicate Gray's experiment.

In fact, the dispute between Renfrew and most of our community has been resolved in favour of the steppe hypothesis, but neither by archaeologists nor by linguists; the crucial evidence is ancient DNA evidence. Haak et al. 2015 demonstrated that there was a major population incursion from the steppes into Europe in the middle of the third millennium BCE – more or less exactly as the steppe hypothesis had posited – and that the distribution of steppe DNA correlates well with later populations known to have spoken Indo-European languages (see especially Mallory 1989). Those findings are irreconcilably inconsistent with Renfrew's scenario, according to which, Indo-European languages should have spread first from Anatolia to the Mediterranean lands and from there to northern Europe. That illustrates the most important contention of this chapter: that information from all disciplines must be used, since any one source of information is inconclusive.

4.6 Conclusion

The general conclusion of this chapter is neither sweeping nor startling. We should use computational cladistics for what it's worth, but we need to be aware that its worth is limited. The general rule about extrapolating into the unobserved past still applies: results are comparatively secure when *different* lines of evidence converge on the *same* result. Computational cladistics yields only one line of evidence; therefore, it must be used in conjunction with traditional methods, archaeology, ancient DNA evidence and everything else that might be relevant.

References

- Anthony, David W. 2007. *The horse, the wheel, and language*. Princeton, NJ: Princeton University Press.
- Anthony, David W. & Don Ringe. 2014. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1. 199–219.

- Bergsland, Knut & Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3. 115–53.
- Berwick, Robert. 2015. Darwin's dream. Paper read at the Workshop on Language Variation and Change and Cultural Evolution, University of York, 12 February 2015.
- Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91. 194–244.
- Dobson, Annette J. No date [1970s]. Unrooted trees for numerical taxonomy.
- Dyen, Isidore, Joseph B. Kruskal & Paul Black. 1992. *An Indo-European classification: A lexicostatistical experiment*. Philadelphia, PA: American Philosophical Society.
- Embleton, Sheila M. 1986. *Statistics in historical linguistics*. Bochum: Brockmeyer.
- Evans, Steven N., Don Ringe & Tandy Warnow. 2006. Inference of divergence times as a statistical inverse problem. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 119–29. Cambridge: McDonald Institute for Archaeological Research.
- Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–9.
- Haak, Wolfgang, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522. 207–11.
- Jespersen, Otto. 1909. *A modern English grammar on historical principles*. Vol. 1. Copenhagen: Munksgaard.
- Mallory, James P. 1989. *In search of the Indo-Europeans*. London: Thames and Hudson.
- Meiser, Gerhard. 1986. *Lautgeschichte der umbrischen Sprache*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Nakhleh, Luay, Don Ringe & Tandy Warnow. 2005. Perfect phylogenetic networks: A new methodology for reconstructing the evolutionary history of natural languages. *Language* 81. 382–420.
- Nichols, Johanna & Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2. 760–820.
- Pereltsvaig, Asya & Martin W. Lewis. 2015. *The Indo-European controversy*. Cambridge: Cambridge University Press.
- Renfrew, Colin. 1987. *Archaeology and language: The puzzle of Indo-European origins*. Cambridge: Cambridge University Press.
- Ringe, Don. 2017. *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.
- Ringe, Don & Joseph F. Eska. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Ross, Malcolm. 1997. Social networks and kinds of speech-community event. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language*. Vol. 1. *Theoretical and methodological orientations*, 209–61. London: Routledge.
- Ross, Malcolm. 1998. Sequencing and dating linguistic events in Oceania: The linguistics/archaeology interface. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language*. Vol. 2. *Correlating archaeological and linguistic hypotheses*, 141–73. London: Routledge.

5 Anatolian

Alwin Kloekhorst

5.1 Introduction

The Anatolian branch consists of a group of languages once spoken in ancient Anatolia (modern-day Turkey) and northern Syria, with textual remains dating from the beginning of the second millennium BCE to the second century CE.¹ It is commonly assumed that in the course of the first millennium CE, the entire Anatolian branch became extinct. The attested Anatolian languages are (in chronological order) as follows.²

Kanišite Hittite:³ a dialect of Hittite proper, which is known from hundreds of personal names and a handful of loanwords attested in Old Assyrian texts (clay tablets, written in the Old Assyrian version of the cuneiform script, dating to *c.* 1935–1710 BCE) mostly stemming from Kaniš/Nēša (modern-day Kültepe), Central Anatolia.

Hittite (“Ḫattuša Hittite”):⁴ the main language of the administration of the Hittite kingdom, written in its own version of the cuneiform script, attested in some 30,000 fragments of clay tablets (dating to *c.* 1650–1180 BCE),⁵ especially found in the Hittite capital Ḫattuša (modern-day Boğazkale), but also

¹ The research for this chapter was conducted as part of the NWO-funded research project *Splitting the mother tongue: The position of Anatolian in the dispersal of the Indo-European language family* (NWO-project number 276-70-026) and the project *Multilingualism and minority languages in ancient Europe*, funded by the HERA Joint Research Program “Uses of the past” (Horizon 2020). I would like to thank Xander Vertegaal and Stefan Norbruis for their useful comments on earlier drafts of this chapter.

² Kroonen, Barjamovic & Peyrot (2018) have recently claimed that a number of personal names that are recorded in texts from Ebla, dated to the twenty-fifth–twenty-fourth centuries BCE, belong to one or more languages “that clearly fall within the Anatolian Indo-European family” (2018: 6). However, no detailed analysis of this material is offered, and at present I therefore regard the linguistic status of these names as too uncertain to make any broad claims.

³ See Kloekhorst 2019 for a full account of this language and its attestations.

⁴ The authoritative synchronic grammar of Hittite is Hoffner & Melchert 2008. Synchronic dictionaries are HW² and CHD; etymological dictionaries are HEG, HED, EDHIL. For historical linguistic treatments, see e.g. Melchert 1994; Kimball 1999; EDHIL.

⁵ But see Kloekhorst & Waal 2019, who argue that a few Hittite tablets may stem from the latter half of the eighteenth century BCE.

several other places in Central Anatolia. It is the best attested Anatolian language by far, and therefore the most important witness of this branch.

Palaic:⁶ known from several passages embedded in Old and Middle Hittite texts (sixteenth–fifteenth century BCE), primarily dealing with the cult of the god Zapaṛḫa. It was the language of the land of Palā, situated in the north-west of Central Anatolia. The Palaic corpus is small, and therefore many basic matters regarding grammar and lexicon are unclear.

Cuneiform Luwian (also called Kizzuwatna Luwian):⁷ only known from cultic passages cited in Hittite texts (dating to the sixteenth–fifteenth century BCE). It was certainly spoken in Kizzuwatna (south-east of Central Anatolia) and possibly also in the western part of Anatolia. In Hittite texts from the New Hittite period (fourteenth–thirteenth century BCE), we find many Luwian loanwords, which traditionally were regarded to be Cuneiform Luwian as well but which may be more appropriately regarded as linguistically belonging to Hieroglyphic Luwian.

Hieroglyphic Luwian (also called Empire Luwian / Iron Age Luwian):⁸ closely related to Cuneiform Luwian, written in an indigenous hieroglyphic script (Marazzi 1998) that seems to have been especially designed for this language. Seals containing these hieroglyphs can be dated as far back as the Old Hittite period (c. 1600 BCE), but real texts (mostly inscriptions on rocks and stone steles) date from the thirteenth to the end of the eighth century BCE. The c. thirty texts that date from the last phase of the Hittite Kingdom (so-called Empire period, and therefore “Empire Luwian”) are found all over Anatolia and northern Syria, whereas the c. 230 post-Empire period inscriptions (Iron Age, and therefore “Iron Age Luwian”) are restricted to south-eastern Anatolia and northern Syria, the region of the so-called Neo-Hittite city states. Thanks to a boost in studies of the language since the publication of Hawkins 2000, Hieroglyphic Luwian has become one of the better-known Anatolian languages.

Lydian:⁹ the language of the land of Lydia (central western Anatolia), written in its own version of the Greek alphabet, attested in some 120 texts (the bulk of which are inscriptions on stone steles), dating from the eighth to the third century BCE (with a peak in the fifth–fourth century BCE). Our

⁶ For texts, grammar, vocabulary and historical phonology, see e.g. Carruba 1970; Kammenhuber 1969; Melchert 1994: 190–228.

⁷ Texts are collected in Starke 1985. For grammatical treatments, see Starke 1990; Melchert 2003. For the lexicon, see Melchert 1993. See Yakubovich 2010 for the term “Kizzuwatna Luwian”.

⁸ Texts can be found in Hawkins 2000, see also ACLT. For grammatical treatments, see Melchert 2003; Payne 2010; Yakubovich 2015. A good Hieroglyphic Luwian dictionary is a desideratum: Meriggi 1962 is largely outdated, and the lexical part of ACLT can only be used with caution.

⁹ For texts, grammar and vocabulary, see Gusmani 1964. Historical linguistic treatments can be found in Melchert 1994: 329–83; Gérard 2005. A more general introduction to the Lydians and their language is Payne & Wintjes 2016.

knowledge of Lydian is limited since there are only a few bilingual texts and since its vocabulary is difficult to compare to the lexicon of the other Anatolian languages (see also below, [Section 5.3.3](#)).

Carian:¹⁰ the language of the land of Caria (south-central western Anatolia), written in its own version of the Greek alphabet, attested in some 200 inscriptions from the seventh–fifth century BCE from Egypt (tomb inscriptions from Carian mercenaries living there) and from the fourth–third century BCE from Caria itself. Our knowledge of Carian is very rudimentary: the Carian alphabet was not successfully deciphered until the 1990s, and many inscriptions contain personal names only.

Lycian (also called Lycian A):¹¹ the language of Lycia (south-western Anatolia), written in its own version of the Greek alphabet, in some 150 coin legends and 170 inscriptions on stone, dating to the fifth–fourth century BCE. Our knowledge of Lycian is relatively advanced, partly because of some bilingual texts (including the large trilingual inscription of Letōon) and partly because of its linguistic similarities with the Luwian languages. Nevertheless, many details regarding grammar and lexicon are still unclear.

Milyan (also called Lycian B):¹² attested in two inscriptions from Lycia (fifth century BCE) that are written in the Lycian alphabet. Although the name “Milyan” refers to the region Milyas, situated in the north-east of Lycia, it is unclear where it originates. The two Milyan inscriptions, which both seem to be in verse, are difficult to understand, and our knowledge of Milyan is therefore rudimentary.

Sidetic:¹³ the language of the city of Side (south coast of Anatolia) and its surroundings, written in its own version of the Greek alphabet, attested in some ten inscriptions on coins and stone, dating to the fifth–second century BCE. The number of textual remains is very low, so we only know a few facts about Sidetic grammar and lexicon.

Pisidian:¹⁴ a language attested in a few dozen tomb inscriptions in the Greek alphabet that were found in the eastern part of classical Pisidia (south-west of Central Anatolia), dating to the first–second century CE. The inscriptions contain only personal names, some of which point to an Anatolian character to this language.

¹⁰ See [Adiego 2007](#) for a full discussion of all Carian texts, and the grammar, lexicon and historical linguistic interpretation of the language.

¹¹ For text editions see [Kalinka 1901](#); [Neumann 1979](#); [Laroche 1979](#). The vocabulary is compiled in [Melchert 2004](#) and [Neumann 2007](#). Grammatical treatments and historical linguistic analyses can be found in e.g. [Hajnal 1995](#); [Melchert 1994](#): 282–328; [Melchert 2004](#); [Kloekhorst 2013](#).

¹² [Shevoroshkin 2013](#). The Milyan vocabulary is included in [Melchert 2004](#) and [Neumann 2007](#).

¹³ [Pérez Orozco 2007](#). ¹⁴ [Brixhe 1988](#).

5.2 Evidence for the Anatolian Branch

There is ample evidence to view the Anatolian languages as forming a single branch: they share enough linguistic features to set them apart as a single group vs. the rest of the Indo-European language family, cf. e.g. [Rieken 2017: 299](#). A complicating factor, however, is the Indo-Anatolian hypothesis, which states that Anatolian was the first branch to split off from the Indo-European mother language, after which the remaining language, which was to become the ancestor language of all the non-Anatolian Indo-European languages (“Core Proto-Indo-European”) underwent a set of innovations (see [Section 5.5](#)). Whenever the Anatolian languages show shared features that are different from the other Indo-European languages, we should therefore investigate to what extent these differences are caused by innovations that took place in the prehistory of the Anatolian branch, or by innovations in the prehistory of Core Proto-Indo-European. In the latter case, the Anatolian features may in fact be shared *retentions*, and therefore cannot, strictly speaking, be used in arguing that the Anatolian languages form a single branch. In practice, however, it is not always easy to distinguish between the two.

Another complicating factor is that some of the Anatolian languages have a very limited attestation (especially Sidetic and Pisidian) or are in general poorly understood (Carian, Milyan and, to a lesser extent, Lydian and Palaic). This means that not all features listed below are found in all languages.

The following specific features of the Anatolian languages can be regarded as examples of common innovations that prove the unity of the Anatolian branch and allow for the postulation of an ancestor language, Proto-Anatolian, from which they all derive:

Phonology

- the merger of PIE *mediae* and *aspiratae* into a single series that is called *lenis* (PIE **d*, **d^h* > PANat. **t*/),¹⁵ which is distinct from the so-called *fortis* series, which is the outcome of PIE *tenues* (PIE **t* > PANat. **t*/)
- the operation of Eichner’s lenition rules: (1) pre-PANat. **ṽC:V* > PANat. **ṽCV* and (2) pre-PANat. **ṽ...VC:V* > PANat. **ṽ...VCV*¹⁶

¹⁵ As argued in e.g. [Kloekhorst 2016: 226–8](#), within the glottalic theory this merger may be seen as the result of the development of PIE *mediae*, which can be interpreted as pre-glottalized lenis stops, e.g. PIE **d* = **[t̪]*, into a biphonemic pair of glottal stop + lenis stop: PIE **d* = **[t̪]* > pre-PANat. **[t̪]* = **ʔ/* + **t̪/*. In this way, the oral part of the PIE *mediae* was detached from its glottal part and merged with the PIE *aspiratae*, which in fact were lenis stops (e.g. PIE **d^h* = **[t̪]*), whereas the glottal stop merged with the outcome of PIE **h₁*.

¹⁶ [Eichner 1973: 79, 100 n. 86](#). The two lenition rules can in fact be regarded as a single development, which may be represented as pre-PANat. **ṽ(̣...)VC:V* > PANat. **ṽ(̣...)VCV*, cf. [Adiego 2001; Kloekhorst 2014: 547–87](#).

- PIE accented short **ó* was lengthened to PAnat. long **ō̄*/ (and subsequently caused lenition according to Eichner's first lenition rule)¹⁷
- the PIE cluster **h₂u* yields PAnat. monophonemic **q^w:/*,¹⁸ e.g. PIE **trh₂u(e)nt-* > PAnat. **/t.rq^w:(ə)nt-/* > Hitt. *tarḫuuant-* /tərḫ^w:ənt-/ , CLuw. *tarḫu(ua)nt-* /tərḫ^w:(ə)nt-/ , Lyc. *trqqāt-* /trk^w(a)nt-/ , Car. *trqδ-* /trk^{wnt}-/
- the development of a lateral in the word for 'name': PIE **h₃neh₃mn-* > PAnat. **/lōmn-/* > Hitt. *lāman-*, HLuw. *álanan-*, Lyc. *alāma-*

Morphology

- the creation of an acc.-dat. form **/m:u(-)/* 'me' (vs. PIE **h₁mmé-*)
- the creation of a demonstrative pronoun **/pōpō-/* (from virtual PIE **h₁o-b^hó-*)¹⁹
- the loss of the distinction between present and aorist (the "tezzi-principle")²⁰
- the creation of the *hi*-conjugation (cognate to the PIE perfect)²¹
- the 1pl. ending **-uén(i)/* (cognate to the PIE dual ending **-ué*)²²
- the replacement of the post-consonantal pret.act.3sg. ending **-t* by the middle ending **-to* (> Hitt. *-tta*, CLuw. *-tta*, HLuw. *-ta*, Lyc. *-te*)²³
- the loss of the subjunctive and optative moods.

For other specifically Anatolian features, see [Section 5.5](#), where a list of shared retentions of Anatolian will be presented (as arguments in favour of the Indo-Anatolian hypothesis).

5.3 The Internal Structure of Anatolian

There is some debate on the exact internal subgrouping of the Anatolian branch, although on some aspects there is broad consensus.

5.3.1 The Luwic Branch

There can be no doubt that Cuneiform Luwian, Hieroglyphic Luwian and Lycian form a separate branch, which is commonly called "Luwic". This

¹⁷ Kloekhorst 2014: 439–59.

¹⁸ Kloekhorst 2006: 102; Melchert 2011: 128–9. Cf. Kloekhorst 2018a for the postulation of a labio-uvular stop **q^w:/* for the PAnat. stage.

¹⁹ EDHIL: 192. ²⁰ Malzahn 2010: 267–8.

²¹ E.g. Eichner 1975; Kloekhorst 2018b, contra Jasanoff 2003.

²² Jasanoff 2003: 3; EDHIL: 1001.

²³ The idea that CLuw. *-tta*, HLuw. *-ta* and Lyc. *-te* reflect the middle ending **-to* is generally accepted (e.g. Yoshida 1993), but the origin of Hitt. *-tta* is debated. Some scholars assume that the spelling *°C-ta* can represent *°Ct/* < PIE **°C-t* (e.g. Yoshida 1991: 28); others assume that the *a*-vowel is real and developed as a prop-vowel, i.e. *°C-ta/* < PIE **°C-t* (e.g. Melchert 1994: 175–6, with references); and still others have argued that the *a*-vowel is real but cannot be explained as a prop-vowel, and that Hitt. *-tta* therefore must reflect earlier **-to* (EDHIL: 800–1). If the latter view is correct, the spread of **-to* at the cost of post-consonantal **-t* must have been a common innovation of all Anatolian languages.

means that these languages derive from a “Proto-Luwic” mother language. It is generally assumed that Milyan and Carian belong to this branch too, and also Sidetic and Pisidian are often regarded as possibly Luwic languages (e.g. Melchert 2003: 170–7; Yakubovich 2010: 6; Rieken 2017: 301–3).

5.3.1.1 Shared Innovations of the Luwic Languages The Luwic sub-branch of Anatolian can be defined through the following innovations (although they are not always attested in all languages):

Phonological

- the assibilation of PANat. */k:/ > PLuwic */ts/ > CLuw. *z* /ts/, HLuw. *z* /ts/, Lyc. *s*, Mil. *s*, Car. *s*, Sid. *š* (vs. Hitt. /k:/, Pal. /k:/, Lyd. *k*)
- the weakening of PANat. lenis */k/ > PLuwic */j/ > Ø: e.g. PANat. */kés:r-/ ‘hand’ > CLuw. *īš(ša)ri-*, HLuw. *istri-*, Lyc. *izri-* (vs. Hitt. /k/ and Pal. /k/)
- the weakening of PANat. lenis */kʷ/ > PLuwic */u/: e.g. PANat. */kʷōu-/ ‘cow’ > HLuw. *wawa/i-*, Lyc. *wawa-*, *uwa-* (vs. Hitt. /kʷ/, Pal. /kʷ/, Lyd. *k*)
- the merger of PANat. */e/ and */ō/ into PLuwic */ǃ/ ²⁴ (vs. their retention as separate phonemes in Hittite, Palaic and probably Lydian ²⁵)
- Čop’s Law: PANat. */ĀCV/ > PLuwic */ĀC:/ ²⁶

Morphological

- the large-scale spread of the proterodynamic *i*-stem inflection replacing original consonant stem and *o*-stem inflection (formerly called “*i*-mutation”) ²⁷
- the reshaping of the PANat. nom.pl.c. ending **-es* to PLuwic **-Vns-i* (based on the acc.pl.c. ending **-Vns* < PIE **-Vms* + the original pronominal nom.pl.c. ending **-i* < PIE **-oi*?) > CLuw. *-Vnzi* /-Vntsi/, HLuw. *-V-zi* /-Vntsi/, Lyc. *-i* (< **-insi*), *-ēi* (< **-onsi*), *-āi* (< **-ānsi*), Car. *-š* (?)
- the grammaticalization of the genitival adjective in **-os:o/i-* > CLuw. *-ašša/i-*, HLuw. *-asa/i-*, Lyc. *-ehe/i-*, Mil. *-ehe/i-*, Car. *-š* (?), Sid. *-asV*, Pis. *-s* (?) ²⁸

²⁴ The development of the Carian vowel system is too poorly understood for us to be certain that Carian was part of this development. If it was not, this isogloss should be removed from the inventory.

²⁵ Note that the prehistory of the Lydian vowel system is still relatively unclear.

²⁶ Cf. Kloekhorst 2014: 567–85 for the fact that this law is not only valid in CLuwian (for which it was originally formulated, cf. Čop 1970), but also in HLuwian and Lycian.

²⁷ See Norbruis 2021: 9–50 (adapting Rieken 2005) for a full treatment of the phenomenon that in the prehistory of the Luwic branch the proterodynamic *i*-stem inflection (synchronically characterized by the presence of an *-i-* in the nom.sg./pl.c. and acc.sg./pl.c. cases vs. the absence of *-i-* in all other cases, therefore termed “*i*-mutated”), which it had inherited from PIE, spread widely within the nominal system, first to consonant-stems, and later to **o*-stems (but not to **ā*- and **u*-stems). See below for the fact that Palaic and Lydian also show some cases of this spread.

²⁸ This suffix is attested in Hittite, too, but it has not been grammaticalized as an inflectional morpheme, cf. EDHIL: 216.

- the spread of the pret.act.3sg. ending **-to* to verbal stems ending in a vowel (at the cost of the original ending **-t*)²⁹

Lexical

- PLuwic **mās:Vn-* ‘god’ > CLuw. *māššani-*, HLuw. DEUS-*ni-*, Lyc. *mahana-*, Mil. *masa-*, Car. *mso-*, Sid. *masara-* (vs. PANat. **tieu-* (< PIE **dieu-*) in Hitt. *šiu-*, Lyd. *ciw-*)³⁰
- PANat. **t.rqʷ:(ə)nt-* ‘(one who has / has been) conquered’ develops into the generic name for ‘Storm-god’ in PLuwic, yielding CLuw. *tarḫu(a)nt-*, HLuw. *tarhunt-*, Lyc. *trqqñt-*, Mil. *trqqñt-*, Car. *trqð-*, all ‘Storm-god’ (vs. Hitt. *tarḫuuant-* ‘conquered’ and ^dIŠKUR-*unn(a)-* ‘Storm-god’)³¹

5.3.1.2 Internal Subgrouping of the Luwic Branch The relationships between the three better-known Luwic languages – Cuneiform Luwian, Hieroglyphic Luwian and Lycian – are quite clear. It is generally accepted that Cuneiform Luwian and Hieroglyphic Luwian are closely related, yet distinct, dialects. The relationship between the two cannot have been a matter of one of them deriving from the other (cf. Melchert 2003: 171–2), which means that both must go back to a common ancestor, which may be termed Proto-Luwian.

Lycian is generally recognized as being closely related to the two Luwian languages. Yet, although it was attested almost a millennium after the latter’s first attestations, it was clearly not a direct daughter language of either of them: the Luwian languages show innovations that are not shared by Lycian (e.g. merger of PLuwic **ə/* and **a/* into PLuwian **a/*; replacement of the dat.-loc. pl. ending **-əs/* (< PANat. **-os/*) by **-əns/*³² > PLuwian **-ants/* (CLuw. *-anza* /-ants/, HLuw. *°a-za* /-ants/); fricativization of **q:/* to PLuwian **χ:/* (Kloekhorst 2018a: 73–6)). This means that Lycian stems from a sister language to Proto-Luwian and that both can be regarded as distinct daughters of Proto-Luwic.

Although our knowledge of Milyan is limited, it is usually seen as being closely related to Lycian. This is based on the fact that these two languages have several linguistic features in common, which may be seen as shared innovations that set them apart from Proto-Luwian: PLuwic **-Vs* > Mil., Lyc. *-V* (as in nom.

²⁹ Whereas Hittite still shows the old opposition between post-consonantal **-to* and post-vocalic **-t*, Palaic has also retained post-vocalic **-t* (e.g. pret.act.3sg. *lu-ki-i-it*). The origin of Lyd. pret. act.3sg. *-l* is not fully clear, unfortunately.

³⁰ Eichner (1974: 64) proposed that Luw. **mās:Vn-* ‘god’ derives from a pre-form **meh_{1/3}-(o)s-h₃on-* ‘freien Willen habend, nach eigenem Belieben handelnd’ (~ Lat. *mōs* ‘custom, usage’, which Eichner translates as “Wille”). An alternative may be to derive **mās:Vn-* from **meh₁ns-en-*, a derivative of PIE **meh₁ns-* ‘moon’.

³¹ Although it cannot be excluded that already in PANat. **t.rqʷ:(ə)nt-* was the name of the Storm-god, and that Hitt. ^dIŠKUR-*unn-* is an innovation, cf. Kloekhorst 2019: 192.

³² Having taken over the nasal of acc.pl.c. **-ns/* < PIE **-ms*.

sg.c.) (vs. PLuwian **-Vs*); PLuwic **-Vn* > Mil., Lyc. *-Ṽ* (as in acc.sg.c.) (vs. PLuwian **-Vn*); *a*-umlaut (e.g. Mil. nom.-acc.pl.n. *uwadra* vs. *uwedr(i)-*, or nom.-acc.pl.n. *χuzruwāta* vs. acc.pl.c. *χuzruwētiz*); syncope in the ethnicon suffix Mil. *-wāni-* and Lyc. *-ñni-* < **-wn:i-* (vs. PLuwian **-wan:i-*); fronting of PLuwic **/k^w/* before a front vowel in rel.pron. **/k^w:i-/* > Lyc. *ti-*, Mil. *ki-/ci-/* (vs. PLuwian **k^wi-*). A shared lexical innovation may be Mil. *kibe* ~ Lyc. *tibe* ‘or’.

The position of Carian, Sidetic and Pisidian is less clear, since the number of possible isoglosses is very low. In the case of Carian, Adiego (2007: 347) states that “a meaningful isogloss shared by Carian and Milyan is the copulative conjunction Car. *sb*, Mil. *sebe* ‘and’”, which contrasts with Lyc. *se* ‘and’. One may add Car. *mso-* ~ Mil. *masa-* vs. Lyc. *mahana-* ‘god’. In the case of Sidetic, the dat.pl. ending *-a* (in *masara* ‘to the gods’), which must reflect PANat. **/-os/*, shows that this language does not belong to the Luwian subgroup (which rather shows the dat.pl. ending **/-ants/*). Furthermore, this ending shows that Sidetic, just like Lycian and Milyan, has undergone the development **-Vs* > *-V*, which may be seen as a shared innovation. On the basis of the Sidetic conjunction *śa* ‘and’, we may assume a closer affinity with Lycian, which has *se* ‘and’ (vs. Mil. *sebe* and Car. *sb*). In the case of Pisidian, a closer affinity with the Lyco-Carian subgroup may be seen from the nom.sg.c. ending *-V*, which then corresponds to Lyc. *-V*, Mil. *-V*, Car. *-Ø* < PLuwic **-Vs* (vs. CLuw. *-Vš* and HLuw. *-Vs*). The exact position of Pisidian within this group must remain undetermined, however.

All in all, the tree of the Luwic sub-branch may be envisaged as in Figure 5.1.

5.3.1.3 Dating Proto-Luwic The Luwic branch seems to have been relatively shallow. As mentioned above, the linguistic difference between

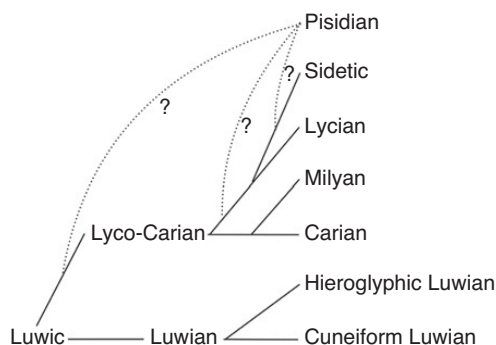


Figure 5.1 The Luwic sub-branch of Anatolian

Cuneiform Luwian and Hieroglyphic Luwian is minimal, and we may therefore date their pre-stage, Proto-Luwian, to not much more than a handful of generations before the oldest attested Cuneiform Luwian texts (sixteenth century BCE), i.e. to *c.* the eighteenth century BCE. In the same vein, the difference between the Lyco-Carian branch and Proto-Luwian seems to have been relatively small, so we may assume that Proto-Luwic preceded Proto-Luwian by no more than two or three centuries. We can thus approximately date this stage to the twenty-first–twentieth century BCE.

5.3.2 *The Position of Palaic*

Since our knowledge of Palaic is limited, it is not easy to determine its position within the Anatolian language family with certainty. Moreover, as Carruba (1970: 4) and Melchert (2003: 269) show, Palaic shares linguistic features both with Hittite and with the Luwic languages, adding to the difficulty. Nevertheless, Oettinger (1978) gives several arguments that would indicate that Palaic is more closely related to the Luwic languages than to Hittite and Lydian. According to Rieken (2017: 303), however, “none of the isoglosses suggested so far [i.e., by Oettinger and others – AK] involve newly created morphology. In each case, the change consists of a choice among several inherited morphemes or a shift of a category’s function, mostly extending it”, and she therefore remains agnostic about the position of Palaic. To my mind, this is too negative a view: there certainly are some features that in fact can be used for judging its place in the Anatolian tree.

- The Palaic dative of the 3sg. enclitic pronoun, =*tu* ‘to him/her’, is identical to CLuw. =*tu* and HLuw. =*du* / =*ru*, but distinct from Hitt. =*šše* (later =*šši*) and Lyd. =*mλ*. Oettinger (1978: 78–9) convincingly argues that this =*tu* originally was the dative form of the 2sg. enclitic pronoun, which was extended to the 3rd person. This non-trivial development was thus a shared innovation of Palaic and the Luwian languages.³³
- In Palaic, Proto-Anatolian lenis */k^w/ (< PIE *g^{w(h)}) is weakened to /x^w/ or /χ^w/ in *aḫuūānti* ‘they drink’ < *h₁g^{w(h)}ēnti. This fricativization may be seen as a first step towards the full weakening that is found in Luwic, where PAnat. */k^w/ > *μ.³⁴
- According to Starke (1990: 71–5), Palaic shows some instances of “*i*-mutation”, indicating a connection with the Luwic branch. Since it has now become clear that the “*i*-mutation” inflection in fact goes back to a normal

³³ Note that the corresponding Lycian morpheme is =*i(je)*, which then must be a later innovation through analogy after the nominal dat.sg. ending -*i(je)*.

³⁴ If Watkins’ suggestion (*apud* Melchert 1990: 207) that Pal. *kuyani-* means ‘womanly’ (i.e. from the PIE stem *g^wen-h₂-) is correct, it would show that this weakening did not take place in word-initial position.

PIE proterodynamic *i*-stem inflection (cf. footnote 27), the mere existence of this type in Palaic is not remarkable per se. However, as noticed by Starke, in Palaic the “*i*-mutated” inflection also seems to be found in original consonant-stems (e.g. ^d*ilaliant(i)-*). This implies a secondary spread that is comparable to the one found in the Luwic branch, and which may then be viewed as a shared innovation. Nevertheless, the fact that our evidence for “*i*-mutated” stems in Palaic is scanty shows that this spread certainly had not yet taken place on such a large scale as in the Luwic languages.

- In Palaic, the pret.act.3pl. ending is *-(a)nta*, which matches Luwic **-Vntā* (CLuw. *-anta*, HLuw. *-anta*, Lyc. *-ṽte*),³⁵ but contrasts with Hitt. *-er* and Lyd. *-rs* / *-riš*. Since **-Vntā* is generally regarded as deriving from the PIE 3pl. middle ending **-ento*, it may be possible to see the transfer of this ending to the pret.3pl. of the active as a common innovation of Palaic and Luwic.³⁶
- In Palaic, the only attested pret.act.1sg. ending is *-(h)ha*, which reflects PANat. **/-q:a/* < PIE **-h₂e*, and thus originally belonged to the *hi*-conjugation. Since it is also found in the form *aniēhha* ‘I did’ (thus Carruba 1970: 50), which was probably originally *mi*-conjugating, it seems that in Palaic the pret.act.1sg. *hi*-ending *-(h)ha* has fully ousted the corresponding *mi*-ending **-m* (attested in Hitt. *-un*, *-nun* and Lyd. *-v*). The same development took place in Luwic, where pret.act.1sg. **-q(:)a* (CLuw. *-(h)ha*, HLuw. *-ha*, Lyc. *-χα*, *-ga*) has fully ousted **-m* as well. We may thus assume that Palaic and Luwic shared this innovation.³⁷

Although the material is scanty and the number of arguments low, it does seem safe to conclude that Palaic shares some innovations with the Luwic branch. Nevertheless, it is clear that Palaic cannot be regarded as a proper Luwic language: for instance, it does not show assibilation of PANat. **/k:/* (which rather yielded Pal. *k*; Melchert 1994: 210), and it does not show a nom.pl.c. ending **-Vnsi* (but rather *-aš* and *-eš*). We should therefore assume that Luwic and Palaic are related on a higher node, which may be termed Luwo-Palaic.³⁸

³⁵ Yoshida’s scenario, by which the Palaic ending *-(a)nta* has a different origin from the Luwic ending **-Vntā* < PIE **-Vnto* (Yoshida 1991: 370–1), seems too complicated to me.

³⁶ We may also assume that this transfer took place as early as in pre-Proto-Anatolian times, and in fact consisted of the replacement of the original pret.act.3pl. ending **-Vnt* < PIE *-(e)nt* by its middle variant **-Vnto* < PIE *-(e)nto* in a reaction to the loss of word-final **-t*, just as pret.act.3sg. **-t* was for the same reason replaced by middle **-to* (cf. Section 5.2). This would fit the fact that a pret.act.3pl. ending **-an* cannot be reconstructed for Proto-Anatolian (contra Yoshida 1991). If correct, we have to assume that Proto-Anatolian, next to *hi*-conjugated pret.act.3pl. **-ēr* / **-rs* (with **-rs* being the original zero-grade variant of **-ēr*?), possessed the *mi*-conjugated ending **-(V)nto*, and that in all branches only one of these endings survived. Hittite generalized the ending **-ēr*, Lydian the ending **-rs*, and Palaic and Luwic the ending **-(V)nto*. This spread of the *mi*-conjugated ending **-(V)nto* at the cost of *hi*-conjugated **-ēr* / **-rs* may then be seen as a shared innovation of Palaic and Luwic.

³⁷ Yakubovich (2010: 6) cites this isogloss as the defining feature of the Luwo-Palaic subgroup.

³⁸ Thus also Oettinger 1978: 92; Yakubovich 2010: 6. See now also Giusfredi 2020: 18–19.

5.3.3 The Position of Lydian

The exact position of Lydian is widely debated, which is due to the fact that this language is poorly understood: only a few Lydian words can be securely translated, making it difficult to establish etymologies and thus sound correspondences with the other Anatolian languages. Nevertheless, there seems to be more and more consensus that Lydian, too, was related to the Luwic sub-branch, since the two share some isoglosses:

- “*i*-mutation” (cf. [Sasseville 2017](#)): since the “*i*-mutation” inflection reflects the normal PIE proterodynamic *i*-stem inflection ([Norbruis 2021](#): 9–50; see also footnote 27), its presence in Lydian is not remarkable per se. However, its presence in nouns like *sfardēt(i)-* ‘Sardian’ (nom.sg.c. *sfardētīš* vs. dat.pl. *sfardētav*), which originally was probably an **-nt*-stem, implies the spread of the “*i*-mutation” inflection at the cost of the consonant-stem inflection, which would be an innovation shared with Luwic (and Palaic).
- Lydian pres.act.1sg. *-u/-w* is identical to PLuwic **-ū* (CLuw. *-ui*, HLuw. *-wi*, Lyc. *-u*),³⁹ which contrasts with Hitt. pres.act.1sg. *-mi* and *-hhi* (unfortunately, in Palaic no pres.act.1sg. forms are attested). However, if **-ū* indeed goes back to the PIA thematic pres.1sg. ending **-oH* ([Kloekhorst 2013](#): 146), the ending is not the result of an innovation. Nevertheless, the fact that both in Luwic and in Lydian (as far as we can tell) **-ū* < **-oH* ousted the athematic *mi*-conjugation ending **-mi* and the *hi*-conjugation ending **-h₂e-i* (which were retained in Hittite, where a newly created **-o-mi* ousted original **-oH*) can be seen as a common innovation.⁴⁰
- Lydian *-cuwe-* ‘to erect(?)’ is regarded by [Oettinger \(1978: 89\)](#) and [Melchert \(2003: 269\)](#) as cognate to CLuw. *tūua-* ‘to place’, HLuw. *tuwa-ⁱ* ‘to place’ and Lyc. *tuwe-* ‘to place (upright)’, which all reflect a stem **tuūV-*. Although there are different views on the exact origin of this formation, it is mostly seen as an innovation, which then must have been shared by Lydian and the Luwic languages.⁴¹

³⁹ As Stefan Norbruis and Oscar Billing (pers. comm.) have pointed out to me, since Lycian does not show a general loss of word-final **-i*, Lyc. *-u* is better derived from PLuw. **-ū* than from **-ūi* or **-ui*. This means that we have to assume that in the Luwian languages the original ending **-ū* was secondarily extended with the present marker **-i*.

⁴⁰ [Yakubovich \(2010: 6\)](#) cites this isogloss as the defining feature of the “Non-Hittite” subgroup.

⁴¹ According to [Oettinger \(1978: 89\)](#), the stem *tuūV-* is based on a false segmentation of the pres.1pl. form **tuūan(i)* < **(d^he-)d^hh₁-uēne(-i)* of the verbal root **d^heh₁-* ‘to put’. [Frotscher \(2012\)](#) argues that **tuūV-* derives from earlier **d^hh₁-oi-*, the stem that is found in Hitt. *dai-ⁱ* / *ti-* ‘to put’, also derived from PIE **d^heh₁-*. And [Melchert \(2004: 74\)](#) rather derives **tuūV-* from a stem **(s)teh₂w-*, ultimately belonging to PIE **steh₂-* ‘to stand’. Since in all languages **tuūV-* means something like ‘to erect’, a connection with PIE **steh₂-* may indeed be more attractive than a connection with **d^heh₁-*. Nevertheless, in all analyses the stem **tuūV-* is to be viewed as an innovation.

Note that we are not necessarily dealing with a shared innovation in all cases in which Lydian coincides with Luwic:

- Lyd. *taada-* ‘father’ < **tóto-* is cognate with PLuwic. **tóti-* (CLuw. *tāti-*, HLuw. *tati-*, Lyc. *tedi-*, Car. *ted*), which differs from Hitt. *atta-* and Pal. *pāpa-* ‘father’. However, since it cannot be excluded that **tóto-* is the Proto-Anatolian form, whereas Hitt. *atta-* and Pal. *pāpa-* are innovations, this isogloss between Lydian and Luwic (see below for the difference in “*i*-mutation”) could in principle represent a shared retention and is therefore non-probative.
- Lydian has a 1sg. reflexive particle =*m*, which is identical to PLuwic *=*mi* (CLuw. =*mi*(?), HLuw. =*mi*), but contrasts with Hittite, which uses =*z(a)* < *=*ti* in this function (no attestations known for Palaic). Since it cannot be excluded that Lydian and Luwic reflect the Proto-Anatolian situation, whereas Hittite may have undergone an innovation, this isogloss may represent a shared retention and therefore is non-probative.

Moreover, there are also some Luwic isoglosses in which Lydian clearly does not participate:

- Panat. lenis /*k^w*/ > Lyd. *k* in *kāna-* ‘woman’ < **g^woneh₂-* (whereas in PLuwic, Panat. */*k^w*/ is weakened to **u*, e.g. **g^woneh₂-* > CLuw. *uāna-*)
- Lyd. *ciw-* ‘god’ < Panat. */*tieu-* < PIE **dieu-* (vs. PLuwic **mās:Vn-* ‘god’)
- Lyd. *a*-stem noun *taada-* ‘father’ (vs. PLuwic “*i*-mutated” **tóti-*, see the forms cited above)

I am therefore reluctant to view Lydian as a proper Luwic language; rather, I assume that both Lydian and Proto-Luwic derive from an earlier node. In order to establish the position of this node vis-à-vis the Luwo-Palaic node as assumed above, the following arguments can be used:

- The Lydian dat.sg. form of the 3rd person enclitic pronoun, =*mλ* ‘to him/her’, can be derived from *=*smei* / *=*smoi* (Kloekhorst 2012: 169), which to my mind is an archaic morpheme (cognate with the PIE element *-*sm-* as found in, e.g., the Skt. pronominal stem *tasm-*). Lydian thus did not participate in the Luwo-Palaic innovation by which the original dat.sg. of the 2nd person enclitic pronoun, *=*tu*, was extended to the 3rd person.⁴²
- The Lydian pret.act.1sg. ending is -*v*, which reflects the Panat. *mi*-conjugation ending *-*m*. Since in Lydian no trace of the corresponding Panat. *hi*-conjugation ending *-*q(·)a* (< PIE *-*h₂e*) is found, we may assume that *-*m* > Lyd. -*v* had been generalized at the cost of *-*q(·)a*. This would then be a reverse development to the generalization of the *hi*-conjugation ending

⁴² It cannot be excluded, however, that Lyd. =*mλ* < *=*smVī* received its -*m*- from the corresponding dat.pl. form *=*smos* (Hitt. =*šmaš*, CLuw. =*mmaš*, Lyd. =*ms*) and originally was *=*soi*, thus being directly cognate with Hitt. =*šše*. If this is the case, Lydian would still show an archaic morpheme vis-à-vis the innovated *=*tu* of Luwo-Palaic, which would indicate that Lydian should derive from a higher node.

*-q(:)a at the cost of *-m that took place in Palaic and Luwic, and which was mentioned above as a possible shared innovation between these latter two branches.

It is for these reasons that I assume that the “Luwo-Lyidian” node must be placed higher up the family tree than Luwo-Palaic (thus also Oettinger 1978: 92; Yakubovich 2010: 6).

5.3.4 *The Position of Hittite*

Hittite proper (“Ḫattuša Hittite”) knew a sister dialect, Kanišite Hittite, that is very similar to it but in some points does deviate (Kloekhorst 2019). This calls for the postulation of a Proto-Hittite ancestor language that may have been spoken only a few generations before the oldest attestations of Kanišite Hittite (twentieth century BCE), i.e. around 2100 BCE.

As has become clear in the sections above, there are no clear linguistic innovations that Hittite shares with any of the other Anatolian languages.⁴³ As Rieken puts it, Hittite is “notorious for [its] conservatism” (2017: 303). However, this does not mean that Hittite can be directly equated with Proto-Anatolian: Hittite, too, has undergone its specific innovations (e.g. the assibilation of dental stop + *i; the almost complete elimination of paradigmatic alternations between fortis and lenis stops; the reshaping of some verbal endings; the transfer of many *mi*-verbs to the *hi*-conjugation (Norbruin 2021: 131–207); the spread of the *n*-suffix in the word for ‘earth’; etc.).

5.3.5 *Dating Proto-Anatolian*

Although it is difficult to say anything certain about the absolute dating of reconstructed ancestor languages, in the case of Proto-Anatolian we have seen that its two best-known branches, Luwic and Hittite, have proto-languages that are roughly contemporaneous: Proto-Luwic can be approximately dated to the twenty-first–twentieth century BCE, and Proto-Hittite to *c.* 2100 BCE. The difference between the two is quite sizable, and elsewhere (Kloekhorst in press) I have therefore argued that they may have been a millennium apart from each other, which would mean that Proto-Anatolian started to diverge sometime around the thirty-first century BCE.

⁴³ At first sight, the fact that both Lyd. *ciw*- and Hitt. *šiu*- < PANat. */tieu-/ ‘god’ show assibilation / palatalization of the word-initial */t/ may be seen as a shared innovation between these two languages. However, since Lydian shows other features that it shares with Luwo-Palaic, we have to assume that the assibilation in the word for ‘god’ is a parallel, not shared, innovation in these languages.

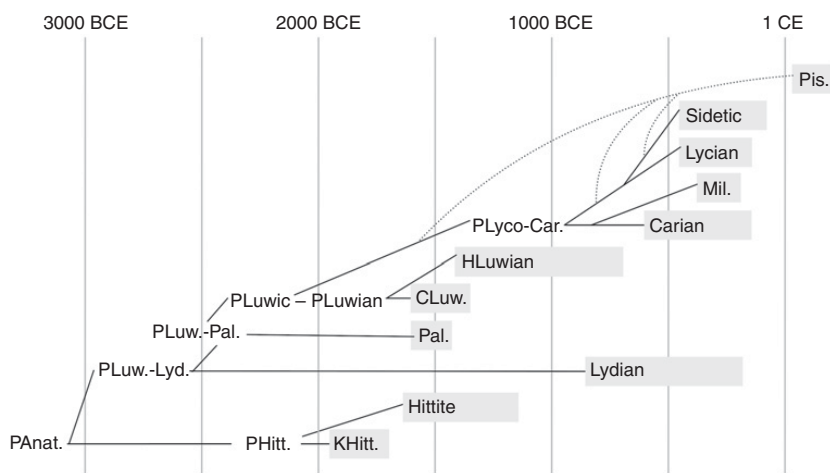


Figure 5.2 Tree model of the Anatolian languages

5.3.6 The Dialectal Make-Up of Anatolian

Taking all these points into account, we arrive at a tree model of the Anatolian branch as shown in Figure 5.2.⁴⁴

5.4 The Relationship of Anatolian to the Other Branches

In 1994, Puhvel argued that Anatolian shares many linguistic features with Celtic, Germanic, Italic, Tocharian and, to a lesser extent, Greek, which would point to a genetic relationship between Anatolian and these “western” branches. For instance, Hitt. *išpant*⁻ⁱ ‘to libate’ matches Lat. *spondeō* and Gr. *σπένδω* but has no cognates anywhere else. However, as Melchert (2016: 300) rightly states, all such cases “can be interpreted as common retentions that just happen to be preserved in Anatolian and the western dialects” and therefore “simply are not probative” for determining the position of Anatolian in the Indo-European family tree: only secured common innovations can be used to this end. Melchert himself thinks that such common innovations between Anatolian and “western” languages may indeed exist, but, to his mind, they would rather prove “post-divergence contact between Anatolian and the western dialects” (2016: 300), and thus have no bearing on the *genealogical* position of Anatolian. Although space limitations do not allow me to examine

⁴⁴ This tree largely coincides with the trees given by Oettinger (1978: 92) and Yakubovich (2010: 6).

the four examples treated by Melchert (2016), it is quite clear that none of them can withstand scrutiny. There is thus no reason to assume that Anatolian shares any innovations, either contact-induced or caused by a genetic relationship, with “western” Indo-European languages or, for that matter, with any of the other Indo-European languages.

5.5 The Position of Anatolian

A hotly debated issue with regard to Anatolian is the so-called Indo-Anatolian hypothesis (also “Indo-Hittite hypothesis”), which states that Anatolian was the first branch to split off from the Indo-European mother language (which then may be called “Proto-Indo-Anatolian” or “Proto-Indo-Hittite”), after which the remaining language, which was to become the mother language of all the non-Anatolian Indo-European languages (and which may be called “Core Proto-Indo-European”, “Nuclear Proto-Indo-European”, “Classical Proto-Indo-European” or similar) underwent a set of innovations. There is some debate on whether this hypothesis is valid at all, and if so, how large the gap is between the moment Anatolian split off and the time that the first split within Core Proto-Indo-European (CPIE) took place (which is usually thought to have been the split-off of Tocharian, see Chapter 6). Some scholars do not think that there is enough evidence for assuming an early split-off of Anatolian at all (Rieken 2009; Adiego 2016); others think that there may have been an early split, but that the gap between Anatolian and the next split is relatively modest (Eichner 2015; Melchert in press), whereas still others think that the gap is sufficiently large for the Proto-Indo-Anatolian ancestor language to be substantially different from Core Proto-Indo-European (Kloekhorst 2008: 7–11; Oettinger 2014; Kloekhorst & Pronk 2019).

The validity of the Indo-Anatolian hypothesis can only be proven if enough secured shared innovations of the non-Anatolian languages can be found. In Kloekhorst & Pronk 2019: 3–6, a total of thirty-four linguistic features are listed in which Anatolian deviates from the other Indo-European languages, and which are presented as possible cases in which Anatolian has retained the original state of affairs, whereas the other Indo-European languages have undergone a common innovation (with twenty-three examples classified as “good candidates”, and eleven as “less forceful” but “promising” ones). These include:

Semantic Innovations

- The Hittite participle suffix *-ant-* forms both active and passive participles, whereas in CPIE the suffix **-e/ont-* only forms active participles: narrowing of the function of **-e/ont-* in CPIE (Oettinger 2014: 156–7).

- The Hitt. active verb $\bar{e}\check{s}^{-zi} < *h_1es^{-ti}$ means ‘to sit’ next to its middle counterpart $e\check{s}^{-a(ri)} < *h_1e-h_1s^{-o}$, which means ‘to sit down’, whereas in C PIE the middle verb $*h_1e-h_1s^{-to}$ means ‘to sit’ next to the verbal root $*sed-$ ‘to sit down’: expansion of the meaning of $*h_1e-h_1s-$ from ‘to sit down’ to ‘to sit’, with replacement of $*h_1e-h_1s-$ ‘to sit down’ by $*sed-$ (Norbruis 2021: 235–41).
- Hitt. $\check{h}arra^{-i} < *h_2erh_3-$ means ‘to grind, crush’, whereas C PIE $*h_2erh_3-$ means ‘to plough’: semantic specialisation in C PIE (Kloekhorst 2008: 9).
- Hitt. $mer- < *mer-$ means ‘to disappear’, whereas C PIE $*mer-$ means ‘to die’: semantic shift, through euphemism, in C PIE (Kloekhorst 2008: 8).

Morphological Innovations

- Anatolian has two genders (common/neuter), whereas C PIE has three genders (m./f./n.): creation of the feminine gender in C PIE (e.g. Melchert *in press*).
- Anatolian has nom. $*ti(H)$, obl. $*tu-$ ‘you (sg.)’, whereas C PIE has nom. $*tuH$, obl. $*tu-$: spread of obl. stem $*tu-$ to the nominative in C PIE (Kloekhorst 2008: 8–9).
- Anat. $*h_1ek_u-$ vs. C PIE $*h_1ek_u-o-$ ‘horse’: thematization in C PIE (Kloekhorst 2008: 10).
- Hitt. $\check{h}uuant- < *h_2uh_1-ent-$ vs. C PIE $*h_2ueh_1nt-o-$ ‘wind’: thematization in C PIE (Eichner 2015: 17–18).

Sound Changes

- Anat. $*h_2 = *[q:]$ and $*h_3 = *[q:w]$ vs. C PIE $*h_2 = *[h]$ or $*[ʕ]$ and $*h_3 = *[h^w]$ or $*[ʕ^w]$: fricativization of uvular stops in C PIE (Kloekhorst 2018a).
- Hitt. $amm- < *h_1mm-$ (< pre-PIE $*h_1mn-$) vs. C PIE $*h_1m-$ ‘me’: degemination of $*mm$ to $*m$ in C PIE (Kloekhorst 2008: 111 n. 234).

Although it is certainly possible that not all of the arguments listed in Kloekhorst & Pronk 2019 will eventually become generally accepted, it seems very unlikely that they will all be refuted, and the Indo-Anatolian hypothesis can thus be regarded as virtually proven. Moreover, since the number of arguments listed is relatively large and some of them concern significant structural innovations (especially the rise of the feminine gender in C PIE, including the creation of the accompanying morphology), it has been argued that the temporal gap between the Anatolian split and the subsequent Tocharian split (cf. Chapter 6) may have been in the range of 800–1000 years. With the Tocharian split commencing around 3400–3300 BCE, the Anatolian split may be dated to the period between 4400–4100 BCE. If Proto-Anatolian indeed first broke up into its daughter languages around the thirty-first century BCE (see Section 5.3.5), it would mean that it had some 1,300–1000 years to undergo the specific innovations that define Anatolian as a separate branch (see Section 5.2). Since these innovations include some large restructurings of especially the verbal system (loss of the subjunctive and optative mood, merger

of the present and aorist aspects, creation of the *hi*-conjugation on the basis of the PIE perfect), such a time span would certainly be fitting.

References

- ACLT = Annotated Corpus of Luwian Texts. <http://web-corpora.net/LuwianCorpus>.
- Adiego, Ignasi-Xavier. 2001. Lenición y acento en protoanatólico. In Onofrio Carruba & Wolfgang Meid (eds.), *Anatolisch und Indogermanisch*, 11–18, Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Adiego, Ignasi-Xavier. 2007. *The Carian language*. Leiden: Brill.
- Adiego, Ignasi-Xavier. 2016. Anatolian languages and Proto-Indo-European, *Veleia* 33, 49–64.
- Brixhe, Claude. 1988. La langue des inscriptions épichoriques de Pisidie. In Yoël L. Arbeitman (ed.), *A linguistic happening in memory of Ben Schwartz*, 131–55. Louvain-la-Neuve: Peeters.
- Carruba, Onofrio. 1970. *Das Palaische: Texte, Grammatik, Lexikon*. Wiesbaden: Harrassowitz.
- CHD = Hans G. Güterbock, Harry A. Hoffner & Theo P. J. van den Hout (eds.). 1983–. *The Hittite dictionary of the Oriental Institute of the University of Chicago*. Chicago, IL: The Oriental Institute of the University of Chicago.
- Čop, Bojan. 1970. Eine luwische orthographische-phonetische Regel. *Indogermanische Forschungen* 75, 85–96.
- EDHIL = Alwin Kloekhorst. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- Eichner, Heiner. 1973. Die Etymologie von heth. *mehur*. *Münchener Studien zur Sprachwissenschaft* 31, 53–107.
- Eichner, Heiner. 1974. *Untersuchungen zur hethitischen Deklination*. Inaugural-Dissertation (Teildruck), Erlangen.
- Eichner, Heiner. 1975. Die Vorgeschichte des hethitischen Verbalsystems. In Helmut Rix (ed.), *Flexion und Wortbildung*, 71–103. Wiesbaden: Reichert.
- Eichner, Heiner. 2015. Das Anatolische in seinem Verhältnis zu anderen Gliedern der indoeuropäischen Sprachfamilie aus aktueller Sicht. In Thomas Krisch & Stefan Niederreiter (eds.), *Diachronie und Sprachvergleich*, 13–26. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Frotscher, Michael. 2012. Die luwischen Entsprechungen der hethitischen Verben des Typs *dāi/tiānzi* und ein neues Lautgesetz urindogermanisch **oi > urluwisch *ue*. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 9, 137–94.
- Gérard, Raphaël. 2005. *Phonétique et morphologie de la langue lydienne*. Louvain-la-Neuve: Peeters.
- Giusfredi, Federico. 2020. *A study in the syntax of the Luwian language*. Heidelberg: Winter.
- Gusmani, Roberto. 1964. *Lydisches Wörterbuch: Mit grammatischer Skizze und Inschriftensammlung*. With Ergänzungsband Lfg. 1 (1980), Lfg. 2 (1982), Lfg. 3 (1986). Heidelberg: Winter.
- Hajnal, Ivo. 1995. *Der lykische Vokalismus*. Graz: Leykam.

- Hawkins, John David. 2000 *Corpus of Hieroglyphic Luwian inscriptions*. Vol. 1. *Inscriptions of the Iron Age*. Berlin: de Gruyter.
- HED = Jaan Puhvel. 1984–. *Hittite etymological dictionary*. Berlin: Mouton.
- HEG = Johann Tischler. 1983–2016. *Hethitisches etymologisches Glossar*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Hoffner, Harry A., Jr. & H. Craig Melchert. 2008. *A grammar of the Hittite language*. Winona Lake, IN: Eisenbrauns.
- HW² = Johannes Friedrich & Annelies Kammenhuber. 1975–. *Hethitisches Wörterbuch*. 2nd ed. Heidelberg: Winter.
- Jasanoff, Jay H. 2003. *Hittite and the Indo-European verb*. Oxford: Oxford University Press.
- Kalinka, Ernestus. 1901. *Tituli Lyciae lingua Lycia conscripti*. Vindobonae: Hölder.
- Kammenhuber, Annelies. 1969. Hethitisch, Palaisch, Luwisch und Hieroglyphenluwisch. In Johannes Friedrich (ed.), *Altkleinasiatische Sprachen*, 119–357. Leiden: Brill.
- Kimball, Sara E. 1999. *Hittite historical phonology*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Kloekhorst, Alwin. 2006. Initial laryngeals in Anatolian. *Historische Sprachforschung* 119. 77–108.
- Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- Kloekhorst, Alwin. 2012. The origin of the Lydian dat.sg. ending *-λ*. *Kadmos* 51. 165–73.
- Kloekhorst, Alwin. 2013. Likijskij jazyk [The Lycian language]. In Y. B. Koryakov & A. A. Kibrik (eds.), *Jazyki mira: Reliktovye indoeuropejskie jazyki Perednej i Central'noj Azii [Languages of the world: Relict Indo-European languages of Western and Central Asia]*, 131–54. Moscow: Academia.
- Kloekhorst, Alwin. 2014. *Accent in Hittite: A study in plene spelling, consonant gradation, clitics, and metrics*. Wiesbaden: Harrassowitz.
- Kloekhorst, Alwin. 2016. The Anatolian stop system and the Indo-Hittite hypothesis. *Indogermanische Forschungen* 121. 213–47.
- Kloekhorst, Alwin. 2018a. Anatolian evidence suggests that the Indo-European laryngeals **h₂* and **h₃* were uvular stops. *Indo-European Linguistics* 6. 69–94.
- Kloekhorst, Alwin. 2018b. The origin of the Hittite *hi*-conjugation. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 89–106. Ann Arbor, MI: Beech Stave.
- Kloekhorst, Alwin. 2019. *Kanišite Hittite: The earliest attested record of Indo-European*. Leiden: Brill.
- Kloekhorst, Alwin. In press. Proto-Indo-Anatolian, the “Anatolian split” and the “Anatolian trek”: A historical linguistic perspective. In Kristian Kristiansen & Guus Kroonen (eds.), *The Indo-European puzzle revisited*. Cambridge: Cambridge University Press.
- Kloekhorst, Alwin & Tijmen Pronk. 2019. Introduction: Reconstructing Proto-Indo-Anatolian and Proto-Indo-Uralic. In Alwin Kloekhorst & Tijmen Pronk (eds.), *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*, 1–14. Leiden: Brill.
- Kloekhorst, Alwin & Willemijn J. I. Waal. 2019. A Hittite scribal tradition predating the tablet collections of Ḫattuša? The origin of the “cushion-shaped” tablets KBo 3.22, KBo 17.21+, KBo 22.1, and KBo 22.2. *Zeitschrift für Assyriologie* 109. 189–203.

- Kroonen, Guus, Gojko Barjamovic & Michaël Peyrot. 2018. Linguistic supplement to Damgaard et al. 2018: Early Indo-European languages, Anatolian, Tocharian and Indo-Iranian. doi:10.5281/zenodo.1240524.
- Laroche Emmanuel. 1979. L'inscription lycienne. In Henri Metzger (ed.), *Fouilles de Xanthos*. Vol. 6. *La stèle trilingue de Létôon*, 49–127. Paris: Klincksieck.
- Malzahn, Melanie. 2010. *The Tocharian verbal system*. Leiden: Brill.
- Marazzi, Massimiliano (ed.). 1998. *Il geroglifico anatolico: Sviluppo della ricerca a venti anni dalla sua "ridecifrazione"*. Napoli: Dipartimento di studi glottoantropologici, Università "La sapienza".
- Melchert, H. Craig. 1990. Adjective stems in *-iyo- in Anatolian. *Historische Sprachforschung* 103. 198–207.
- Melchert, H. Craig. 1993. *Cuneiform Luvian lexicon*. Chapel Hill, NC: n.p.
- Melchert, H. Craig. 1994. *Anatolian historical phonology*. Amsterdam: Rodopi.
- Melchert, H. Craig. 2003. Language. In H. Craig Melchert (ed.), *The Luwians*, 170–210. Leiden: Brill.
- Melchert, H. Craig. 2004. *A dictionary of the Lycian language*. Ann Arbor, MI: Beech Stave.
- Melchert, H. Craig. 2011. The PIE verb for 'to pour' and medial $*h_3$ in Anatolian. In Stephanie Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 22nd UCLA Indo-European Conference*, 127–32. Bremen: Hempen.
- Melchert, H. Craig. 2016. "Western affinities" of Anatolian. In Bjarne S. S. Hansen et al. (eds.), *Etymology and the European lexicon*, 297–305. Wiesbaden: Reichert.
- Melchert, H. Craig. In press. The position of Anatolian. In Michael Weiss & Andrew Garrett (eds.), *Handbook of Indo-European studies*. Oxford: Oxford University Press.
- Meriggi, Piero. 1962. *Hieroglyphisch-hethitisches Glossar*. Wiesbaden: Harrassowitz.
- Neumann, Günter. 1979. *Neufunde lykischer Inschriften seit 1901*. Vienna: Österreichische Akademie der Wissenschaften.
- Neumann, Günter. 2007. *Glossar des Lykischen: Überarbeitet und zum Druck gebracht von Johann Tischler*. Wiesbaden: Harrassowitz.
- Norbruis, Stefan. 2021. *Indo-European origins of Anatolian morphology and semantics: Innovations and archaisms in Hittite, Luwian and Lycian*. Amsterdam: LOT.
- Oettinger, Norbert. 1978. Die Gliederung des anatolischen Sprachgebietes. *Zeitschrift für vergleichende Sprachwissenschaft* 92. 74–92.
- Oettinger, Norbert. 2014. Die Indo-Hittite-Hypothese aus heutiger Sicht. *Münchener Studien zur Sprachwissenschaft* 67. 149–76.
- Payne, Annick. 2010. *Hieroglyphic Luwian: An introduction with original texts*. 2nd rev. ed. Wiesbaden: Harrassowitz.
- Payne, Annick & Jorit Wintjes. 2016. *Lords of Asia Minor: An introduction to the Lydians*. Wiesbaden: Harrassowitz.
- Pérez Orozco, Santiago. 2007. La lengua sidética, Ensayo de síntesis. *Kadmos* 46. 125–42.
- Puhvel, Jaan. 1994. West-Indo-European affinities of Anatolian. In George E. Dunkel et al. (eds.), *Früh-, Mittel-, Spätindogermanisch*, 315–24. Wiesbaden: Reichert.
- Rieken, Elisabeth. 2005. Neues zum Ursprung der anatolischen *i*-Mutation. *Historische Sprachforschung* 118. 48–74.
- Rieken, Elisabeth. 2009. Der Archaismus des Hethitischen: Eine Bestandsaufnahme. *Incontri Linguistici* 32. 37–52.

- Rieken, Elisabeth. 2017. The dialectology of Anatolian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 1, 298–308. Berlin: De Gruyter Mouton.
- Sasseville, David. 2017. The Lydian nominal paradigm of *i*-mutation. *Indo-European Linguistics* 5. 130–46.
- Shevoroshkin, V. V. 2013. Milijskij jazyk [The Milyan language]. In Y. B. Koryakov & A. A. Kibrik (eds.), *Jazyki mira: Reliktovye indoevropskie jazyki Perednej i Central'noj Azii* [Languages of the world: Relict Indo-European languages of Western and Central Asia], 154–66. Moscow: Academia.
- Starke, Frank. 1985. *Die keilschrift-luwischen Texte in Umschrift*. Wiesbaden: Harrassowitz.
- Starke, Frank. 1990. *Untersuchungen zur Stammbildung des keilschrift-luwischen Nomens*. Wiesbaden: Harrassowitz.
- Yakubovich, Ilya. 2010. *Sociolinguistics of the Luvian language*. Leiden: Brill.
- Yakubovich, Ilya. 2015. The Luwian language. In *Oxford handbooks online*. doi:10.1093/oxfordhb/9780199935345.013.18.
- Yoshida, Kazuhiko. 1991. Anatolian verbal endings: The third person plural preterite. *Journal of Indo-European Studies* 19. 359–74.
- Yoshida, Kazuhiko. 1993. Notes on the prehistory of the preterite verbal endings in Anatolian. *Historische Sprachforschung* 106. 26–35.

6 Tocharian

Michaël Peyrot

6.1 Introduction

The Tocharian languages A and B are attested in manuscripts from the northern Tarim Basin, present-day Northwest China. Tocharian B is attested from about the fifth to the tenth centuries of the Common Era. Originally from Kuča, it spread east to Yānqí and Turfan, probably in the late sixth and in the seventh century. In Tocharian B itself, the language is referred to as the language of *kuśi* ‘Kuča’. Tocharian A is attested a little later, from about the seventh to the tenth centuries. It is originally from Yānqí, spread with Tocharian B east to Turfan, but not west to Kuča, and is referred to as the language of *ārśi* ‘Yānqí’. Both languages are written in the Indian Brāhmī script, and the vast majority of the manuscripts are of Buddhist content.

Traces of a third Tocharian language have been claimed to be preserved in the Middle Indic Gāndhārī dialect of Niya in the southern Tarim Basin ([Burrow 1935](#)). This hypothesis has not received wide support and must still be considered very uncertain (see further below in [Section 6.3](#)).¹

6.2 Evidence for the Tocharian Branch

The existence of the Tocharian branch of Indo-European is beyond any doubt. The two languages A and B are closely related and share numerous significant innovations, so it is unnecessary to give a full list here. Some of the more important, branch-defining developments are:

- loss of the threefold Proto-Indo-European distinction between the conventionally termed voiceless, voiced and voiced aspirated stops, i.e. **k̥*, **g*, **gʰ* merged into **k* (on **d*, see below);

¹ I will not discuss in detail a posthumously published proposal by Schmidt (2018: 161–271) to read previously undeciphered manuscript fragments in Formal Kharoṣṭhī as a Tocharian variety from Lōulán. His tentative decipherment is not convincing. Instead, these fragments are probably written in an Iranian language related to Khotanese and Tumšūqese ([Dragoni, Schoubben & Peyrot 2020](#)).

- several mergers and shifts in the vowel system, including loss of vowel length, merger of **i*, **e*, **u* into **ə* (the first two regressively palatalising), shifts of **o* to **e* and of **ā* < **eh₂* to **o*, monophthongisation of **ei* to **i* and of **eu* to **'u*, etc.;
- rise of distinctive and morphological palatalisation, principally through the transformation of the contrast between **o* : **ē* into **e* : **'e* and **∅* : **e* into **ə* : **'ə*;
- loss of word-final **-s*, **-m*, **-n*, **-t* (**-d*), which has led to heavy restructuring of both the nominal and the verbal inflection;
- rise of agglutinative case inflection in the noun, next to agglutinative number inflection in some noun classes;
- almost complete loss of prefixing morphology;
- rise of an intricate system of verbal derivation to form intransitives and transitives or causatives;
- numerous significant innovations in the lexicon.

Even considering the late attestation of the Tocharian branch, the extent of structural change is surprisingly large, and it can be argued that this is partly due to a substrate effect. The loss of the distinction between the so-called voiceless, voiced and voiced aspirated stops, the rise of agglutinative case inflection, and the functions of these case suffixes, which include the perlocative, denoting movement through, along or over something, point to Uralic influence. A pre-Proto-Tocharian phase of the vowel system can be compared more specifically with an early form of Samoyedic. Pronoun suffixes attached to the finite verb denoting the object may be compared with the objective inflection in Uralic (Peyrot 2019a with references; on the vowel system, see Warries *in press*).

It is more difficult to assess the Iranian impact on Tocharian. There has been considerable Iranian influence on the lexicon (Isebaert 1980; Tremblay 2005), but only the oldest layer of borrowings from Old Iranian may possibly be added as a branch-defining feature of Tocharian. The reason is that any feature defining the whole branch should have been acquired before the break-up of unitary Tocharian into Tocharian A and B. This is clearly the case with the structural shift attributed to Uralic above. However, many borrowings from Iranian are to be dated after the break-up and therefore do not define the Tocharian branch as such. Examples of this include borrowings from Bactrian, such as Toch.B *akālk* and Toch.A *ākāl* ‘wish’ from Bactrian *αγαλγο* /*ayalg*/: the *ā* *ā* vocalism of Tocharian A, instead of the *ā* *a* vocalism regular in inherited vocabulary, shows that the word has entered the language later, and the Toch.B and Toch.A forms cannot be reconstructed to a common proto-form. Bactrian influence is therefore to be dated after the split of Proto-Tocharian. The case of borrowings from Old Iranian is different. An example is

Toch.B *perne*, Toch.A *parām* ‘glory’, which allows a Proto-Tocharian reconstruction **perne*, borrowed from Old Iranian **farnah-* (Av. *xʷarənah-*).

Nevertheless, for the Old Iranian layer, the details are not fully clear either. Tocharian B would have preserved a word like **perne* unchanged, and the amount of change in Tocharian A is limited: **e > a* in the first syllable; apocope of **e* in the final syllable; *ä*-epenthesis in the final cluster *-rn*. Since these changes in Tocharian A cannot be dated exactly, it cannot be excluded that **farnah-* was borrowed into Tocharian B and A independently, at an early stage, before the relevant sound changes in Tocharian A occurred but after the break-up of Proto-Tocharian. A reason to consider this more complicated chronology are the sound changes **rn > rr* and **ln > ll* in both languages. Good examples of the former are not found in Tocharian A, but the latter is certain. Since old geminates are generally simplified in Tocharian A, the rise of new geminates from **rn* and **ln* must be dated after the general simplification of geminates. The preservation of *rn* in ‘glory’ thus suggests an early but post-Proto-Tocharian borrowing according to the following relative chronology:

1. break-up of Proto-Tocharian;
 2. degemination in pre-Tocharian A;
 3. assimilation of **rn*, **ln* to *rr*, *ll* (the same change occurred independently in pre-Tocharian B);
 4. borrowing of **farnah-* as **perne* (the same borrowing occurred independently in pre-Tocharian B);
 5. **e > a*, apocope of final **e*, and *ä*-epenthesis to produce Tocharian A *parām*.
- Another indication of this chronology is offered by Toch.B *etswe* ‘mule’, borrowed from Old Iranian **atswa-* ‘horse’ (Av. *aspa-*). Although Toch. B *mättsa-*, Toch.A *nätswā-* ‘starve’ shows that Proto-Tocharian **tsw* has developed to *tsts* in pre-Tocharian B after the break-up of Proto-Tocharian, *etswe* has *tsw* unchanged, suggesting that the borrowing is post-Proto-Tocharian. Old Iranian borrowings can only be taken as a branch-defining feature if the preservation of the cluster *tsw* in Tocharian B, and of the cluster *rn* in both languages, receives an alternative explanation, notably a conditioning of the relevant assimilations, such as a difference in accent.

6.3 The Internal Structure of Tocharian

As Tocharian A cannot be derived from Tocharian B or vice versa, a common ancestor called Proto-Tocharian needs to be reconstructed. For instance, Toch. B *yente* ‘wind’ cannot have yielded Toch.A *want* ‘wind’, and the reverse is also impossible: a preform **wente* is to be posited, with innovations in both languages leading to the attested forms. There is no need to discuss the internal subgrouping of Tocharian, since only one tree is possible. The dating of Proto-Tocharian, the only node in this tree, will be discussed below. Even though

Burrow's hypothesis of a third Tocharian language is too uncertain to be taken into account for inferences on the prehistory of Tocharian, it presents an illustrative case for the methodology of internal subgrouping.

The Gāndhārī words in the documents from Niya for which Burrow (1935) suggests a Tocharian etymology are few, and among these only two are relevant here: *kitsa* 'itsa, a title, and *aṃklatsa*, a type of camel. *kitsa* 'itsa has a very Tocharian-looking structure and has been convincingly connected to Toch.B *ktsaitse* 'old', Toch.A *ktsets* 'perfect' by Burrow, who suggests 'elder' for the Gāndhārī title. Toch.B *ktsaitse* derives from PToch. **kəṛaitʰse* with degemination after a diphthong,² and Toch.A *ktsets* has undergone apocope of final *-e* and monophthongisation of **ai* to *e*; both languages have syncopated the **ə* in the first syllable. Niya *kitsa* 'itsa could derive from Tocharian B as well as pre-Tocharian A or a third branch and is therefore useless for subgrouping. It could reflect an older form of the type **kəṛaitʰse* with *i* for *ə* and the regular Gāndhārī final *-a* for the regular Tocharian final *-e*. The geminate could be simplified or left unwritten. Equally, it could go back to a form of the type Toch.B *ktsaitse*, with *i* -epenthesis in the first syllable. Since Tocharian A is attested from the seventh century onwards, much later than Niya Gāndhārī, which is from the third–fourth centuries, it could also derive from an early form of Tocharian A in which monophthongisation of **ai* to *e* had not yet taken place.

The key form for Burrow's understanding of the internal subgrouping is *aṃklatsa* (1935: 673). According to him, *aṃklatsa* denotes a relatively cheap camel, which may therefore have been untrained. He connects the word to Toch.B *aknātsa*, Toch.A *āknats* 'fool', which is formed with the negative prefix **en-* from the verb **kna-* 'to know': in both languages, the vowel of the prefix has been affected by *a*-umlaut, and its nasal has been lost before the cluster *kn-*. To explain the different cluster *ṃkl* in Niya Gāndhārī, he assumes that it goes back to an earlier form with **nkn* that was dissimilated to *nkl*, written *ṃkl*. Since the first *n* of the cluster is lost in both Tocharian A and B, he concludes that the Tocharian variety he assumes in the Gāndhārī of Niya is of a different branch, and this is the reason why it is often termed "Tocharian C".

Burrow's Tocharian etymology of Niya Gāndhārī *kitsa* 'itsa is attractive, but his explanation of *aṃklatsa* is not convincing in view of the semantic and formal problems. At any rate, this questionable etymology can never alone bear the weight of proving a third branch of Tocharian, the famous "Tocharian C".³

² See Peyrot 2008: 45 on Toch.B *-auñe* < **euññe* with the same degemination. Adams cites the word as *ktsaitstse* (2013: 263), but this form is not attested.

³ Even in the unlikely event that the etymology should be correct nevertheless, it does not necessarily prove the existence of a third branch of Tocharian. Rather than being a shared innovation of Tocharian A and B, the change **nkn* > *kn* may be a parallel development, since there are cases where the nasal is lost in Tocharian A but preserved in Tocharian B (Hilmarsson 1991: 193–8).

Rather, in the light of research by Niels Schoubben, who proposes new and convincing alternative explanations for some other items that Burrow explained as Tocharian (Schoubben 2021), scepticism about Burrow's hypothesis is definitely due.

No absolute date can be given for Proto-Tocharian, by definition the latest phase of unity before the break-up in pre-Tocharian A and pre-Tocharian B. The languages are closely related, but differences are considerable in the lexicon, and most scholars estimate Proto-Tocharian around 500 BCE: some take it to be a little bit earlier, between 1000 BCE and 500 BCE; others a little bit later, between 500 BCE and the beginning of the Common Era (see the useful overview of different estimates in Mallory 2015: 7–8).

It is commonly agreed that the advent of Buddhism was after the break-up, as such basic terms as *dharma* 'law' (Toch.B *pelaikne*, Toch.A *märkampal*) and *karman* 'act, fate' (Toch.B *yāmor*, Toch.A *lyalypu*) are different (Lane 1966). But since Buddhism arrived late in the region, perhaps in the first or second century CE, this gives only an unsurprising *ante quem* date.

Contacts with the Iranian languages Bactrian and Sogdian took place after the split, probably in the early first millennium CE. Contacts with Old Iranian are more interesting: since it can be debated whether they occurred before or after the break-up, they may have to be dated close to that break-up. In the scenario sketched above, they would have occurred soon after it.⁴ However, the Old Iranian loanwords are themselves difficult to date in absolute terms. The archaic appearance of words such as Toch.B *etswe* 'mule' ⇐ OIr. **atswa-* 'horse' (Av. *aspa-*) or Toch.B *waipecce* 'possessions' ⇐ OIr. **hwai-paθya-* (Av. *xʷaēpaiθīia-* 'own') suggests a date in the middle of the first millennium BCE or earlier, but a more precise dating is difficult. I have suggested that these loanwords may be associated with the presence of Andronovo related groups in Northern Xinjiang in the thirteenth–ninth centuries BCE (Peyrot 2018: 280), which would accordingly push the date of Proto-Tocharian towards the beginning of the first millennium BCE. The assumed contacts with Uralic, which may date to around 2500 BCE, in any case took place long before the split, in a pre-Proto-Tocharian phase.

Archaeological evidence on the Tocharians themselves is at present not clear enough (Mallory 2015: 29 and *passim*). It is uncertain whether the Cháwúhūgōu cultural group near Qarašähär (Debaine-Francfort 1989: 183–9), whose different phases together cover almost the entire first millennium BCE, can be identified with early speakers of Tocharian A, or whether the Hälädün cultural group of the early first millennium BCE in and near Kuča

⁴ It is tempting to consider the possibility that the apparently impressive technological advances brought by the Iranians speaking this Old Iranian language were the impetus for the split of Proto-Tocharian. At present no evidence for or against this scenario seems to be available.

([Debaine-Francfort 1988](#): 23) can be identified with early speakers of Tocharian B. Accordingly, archaeological evidence for the date of Proto-Tocharian or the place where it was spoken is presently indirect at best.

6.4 The Relationship of Tocharian to the Other Branches

It is now commonly held that Tocharian has no closer affinity to any other branch of Indo-European.⁵ Proposals for closer affinity have been made but have found little acceptance and concern superficial similarities, such as the spread of the *n*-stems in the nominal inflection, which would be shared with Germanic ([Adams 1988](#): 5), or the endings in *-r* of the middle, suggesting a link with Italo-Celtic (e.g. [Lane 1970](#): 78, who attributes the correspondence to post-Proto-Indo-European contact), and so on. References to and discussion of these and other suggestions can be found in [Hackstein 2005](#) and [Malzahn 2016](#): 281.

Not accepting any of the adduced old comparisons, [Hackstein \(2005\)](#) proposes instead several close matches between Tocharian and other branches in grammaticalisation processes. According to him, the observed grammaticalisation processes are independent and parallel instead of shared, and indicate post-Proto-Indo-European contact. The matches that he proposes are with Latin, Slavic, Gothic, Greek and Armenian. Although the cases discussed are interesting, the large number of languages in the comparison makes it unlikely that the parallelisms are due to early contact. In addition, it is open to debate whether the parallelisms, if correctly identified, are indeed so salient that they cannot have come about completely independently. For instance, the univerbation of interrogative and demonstrative in Toch.B *kuse* ‘who’ < **kʷi* + *so*, in Alb. *kush* ‘who’ < **kʷis* + *so*, and in OCS *kъto* ‘who’ with *-to* from PIE **tod* ([Hackstein 2005](#): 177) has not proceeded in exactly the same way; it probably compensates, at least in part, for the loss of inflection and word weight; and it appears to be a natural process. Toch.B *ʃ* and *ʃpā* ‘and’, which Hackstein derives from **h₁eti* and **h₁eti-h₁epi* respectively, in fact represent one and the same etymon **ʃpə* with simplification of *ʃp* to *ʃ* in classical and late Tocharian B ([Peyrot 2008](#): 68) so that *ʃ* cannot be directly compared with Latin *et* or Gothic *ih* (pace [Hackstein 2005](#): 176).

A different case is presented by matches with Anatolian, of which several have been proposed that appear to be fairly solid: see for instance [Pinault 2006a](#): 93. These must be archaisms, not showing any closer affinity between Anatolian and Tocharian, and are potentially relevant to establish the position of Tocharian in the tree of Indo-European, discussed in the [following section](#).

⁵ The prolonged contact with Iranian and the shorter but dramatic impact of Indic are obviously to be discarded as secondary.

6.5 The Position of Tocharian

Tocharian is often claimed to have been the second branch to split off the Indo-European proto-language: after Anatolian, but before all other attested branches. This hypothesis may be called the “Indo-Tocharian” hypothesis, based on the model of Indo-Anatolian (Peyrot 2019b; see Figure 6.1). “Indo-Anatolian”, equivalent to “Indo-Hittite”, is used here in a technical sense for the highest node in the Indo-European tree, before Anatolian split off as the first branch, a scenario for which the evidence is steadily growing (cf. Kloekhorst & Pronk 2019).⁶ Strikingly, the arguments that have been advanced in support of the “Indo-Tocharian” hypothesis vary considerably: many authors making the same claim do not accept each other’s evidence for their claim. The most comprehensive systematic review is that by Ringe (1991), who finds hardly any evidence for the position of Tocharian in the family tree at all. Other relevant contributions include Lane 1970, Schmidt 1992, Winter 1997, Pinault 2013 and Malzahn 2016.

Below, a selection of arguments will be discussed. In general, it appears that aberrancies of Tocharian are due to innovation, and careful reconstruction tends to bring Tocharian closer to non-Anatolian Indo-European. The Indo-Tocharian

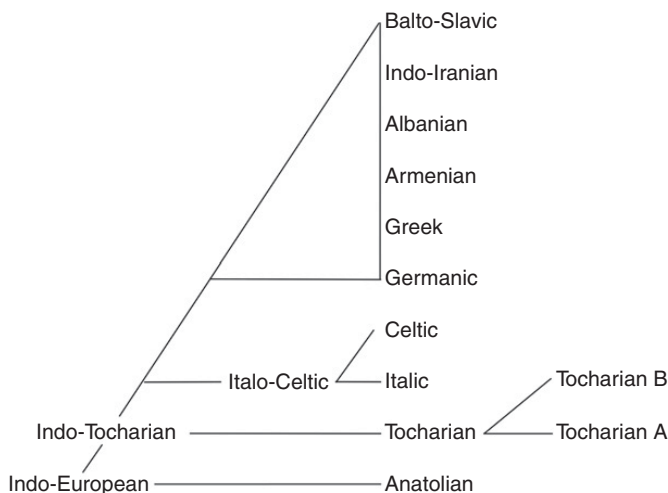


Figure 6.1 The position of Tocharian

⁶ Jasanoff (2017: 233–4) explicitly subscribes to this scenario but rejects the term “Indo-Hittite” because it “acquired tendentious overtones” (p. 233).

hypothesis still seems attractive, but evidence is slim and the difference between Indo-Anatolian and Indo-Tocharian appears to be much larger than that between Indo-Tocharian and the other Indo-European languages. If Indo-Anatolian can be dated to the middle of the fifth millennium BCE, Indo-Tocharian must be much closer to the middle of the fourth millennium. As pointed out to me by Tijmen Pronk, the split-off of the Tocharian branch (Anthony 2007: 305, 307–11; Anthony & Ringe 2015: 208, 211) may be associated with the apparent abandonment of the Caspian steppe in 3500–3400 BCE, probably due to abrupt aridification (Shishlina 2008: 220).

6.5.1 Methodology

In view of the many different arguments that have been proposed for the Indo-Tocharian hypothesis, a brief note on the methodology seems in order.

It is generally agreed that the assumption of an early Tocharian split-off must be based on shared innovations of the other non-Anatolian Indo-European languages. In particular, the branch that split off *after* Tocharian should have shared in such innovations. As the most likely candidate for the branch to have split off third appears to be Italo-Celtic, the supposed shared innovation should ideally be attested in this branch. Conversely, arguments for Indo-Anatolian should be based on shared innovations of non-Anatolian Indo-European, ideally attested also in Tocharian (Peyrot 2019b).

Though clear in theory, in practice finding and defining shared innovations is difficult. There appear to be the following requirements to shared innovations useful for phylogenetic subgrouping:

- *identifiability*: the linguistic element adduced as a shared innovation in the lower node should be clearly identifiable in the higher as well as in the lower node;
- *unidirectionality*: the observed difference with regard to the selected linguistic element should be interpretable as a unidirectional change;
- *salience*: the observed change should be so salient that it is unlikely to have occurred independently in the supposed lower-node branches, in which case it would be a parallel, not a shared innovation.

The requirement of unidirectionality is widely accepted, and discussion tends to focus on the question as to whether a given difference can be interpreted as a unidirectional change, rather than the need of this requirement as such. A case in point is semantic change: phylogenetic arguments based on semantic change are often contested on the grounds that a given semantic difference is not necessarily due to unidirectional change.

The requirement of identifiability, often implicit, may be helpful in discussions about debated phylogenetic arguments based on the loss or addition of features, or on lexical replacement. Arguments based on loss or addition are

notoriously difficult, as for instance with the comparative and superlative suffixes, which are unattested in Anatolian and Tocharian: have they been lost in both branches, or were they added after the Tocharian split-off? Such arguments cannot be applied if the supposedly added feature cannot be identified with any prestige leading to it or if the lost feature has left no trace at all. Arguments based on lexical replacement are weak because the identifiable element would be the meaning, expressed with different etyma in two branches. Meaning is difficult to use as an identifiable element, because several etyma may have similar, overlapping or even identical meanings, and it is therefore difficult to prove that a certain meaning came to be expressed with a different etymon.

The requirement of salience seems so obvious that no further explanation is needed.

6.5.2 Phonology

For our present purposes, phonological evidence appears to be of little relevance in view of the extensive changes in the Tocharian sound system, which are probably due to a Uralic substrate (see Section 6.2). In particular, evidence for the phonetic realisation of the stops in the proto-language has been obscured by this substrate effect. Thus, there is little evidence to establish the position of Tocharian with relation to Kloekhorst's claim that Anatolian preserves an older system of stop distinctions (2016), with classical PIE $*t$, $*d$, $*d^h$ from Proto-Indo-Anatolian $*t_1$, $*t_2$, $*t_3$.

For Tocharian, the developments of $*d$ and $*b^h$ are notable. PIE $*d$ is by default represented with ts and is otherwise often lost, at least before $*i$, $*u$ and $*r$, and so differs from $*t$ and $*d^h$, whose default outcome is Tocharian t .⁷ Thus, even though the exact phonetics remain difficult to establish, $*t$ and $*d^h$ were apparently closer to each other than either of them were to $*d$.⁸ At the same time, $*b^h$ is lost after $*m$, for instance in $*gomb^ho-$ > Toch.B *keme* 'tooth', while $*p$ stays, for instance in $*temp-$ (Lith. *tempiù* 'stretch') > Toch.B *cəmp-* 'be able'.⁹ This suggests that $*b^h$ was weaker than $*p$: it may have been voiced,

⁷ Also, the palatalised reflex of $*d$ is $ś$, while that of $*d^h$ and $*t$ is c .

⁸ Possibly, this distribution also holds for the assibilated variant $-s$ < $*-ti$, $*-d^hi$ (Jasanoff 1987), although good evidence for the development of $*-di$ is thus far lacking.

⁹ With original $*mt$, compare also $*d)kmtóm$ '100' > Toch.B *kante*. Parallel cases with $*d^h$, $*ǵ^h$, $*g^h$ and $*g^{wh}$ are not readily available. Ringe discusses the possibility that the Toch.B subj. stem of *lät-* 'go out' as in the inf. *lantsi* shows /lən-/ < $*h_1lucn)d^h$ (1996: 43). However, he notes that forms with a geminate nn like 1sg. *lannu* 'I will go out' rather suggest an original $*ləntn-$. Indeed, all forms with a nasal in Toch.B can probably be derived from *lənn-* < $*ləntn-$, which arose secondarily through suffixation with $*-nəsk-$ in the present (Peyrot 2013: 446). Toch.B *lank_utse* 'light' < $*h_1leng^{wh}u-$ shows that $*g^{wh}$ was not lost after $*n$. It may be supposed that $*ǵ^h$ and $*g^h$ were not lost after $*n$ either. The reason for this exception could be that there was no corresponding velar nasal phoneme and the velar stop had to remain in order to keep the velar nasal allophone.

fricative or both. It is tempting to compare the typologically common loss of voiced stops after nasals, as in English *lamb* /læm/, and posit the value [b] for $*b^h$, but this is certainly not the only option. Combining the evidence from dentals and labials, it appears that the stop system inherited by Tocharian had strong stops for the conventional voiceless stops like $*t$, weak stops for the conventional voiced aspirated stops like $*d^h$, and a series that was different from both for the conventional voiced stops like $*d$.¹⁰ Although Tocharian offers no direct evidence for the reconstruction of glottalic stops in Proto-Indo-European, the fact that $*d$ has a different reflex from $*t$ and $*d^h$ is neatly compatible with it, since under Kortlandt's glottalic theory (e.g. 1985; 2018a) $*d$ [ʔd] on the one hand is set apart from $*t$ and $*d^h$ on the other.

Nevertheless, the value for the phylogenetic position of Tocharian remains undecided. Since there is strong evidence for $*d = *d$ in classical Indo-European, this feature cannot be used. Further, the position of Tocharian cannot be determined with regard to Kloekhorst's claim that classical PIE $*t$ (perhaps phonetically [t]) < Proto-Indo-Anatolian $*t$: and classical PIE $*d^h$ (perhaps phonetically [d]) < Proto-Indo-Anatolian $*t$, since both phonetic stages are compatible with $*t$ being stronger and $*d^h$ being weaker.¹¹

It has been argued that Tocharian shows consonantal reflexes of PIE $*H$ as *k* (e.g. Winter 1965: 206–10; Schmidt 1988; Kortlandt 2018b). Winter adduces Tocharian A “intrusive *k*” as a consonantal reflex of $*HH$, e.g. gen.pl. *lwākis* to nom.-obl.pl. *lwā* ‘animals’ or perl.pl. *puklākā* to nom.-obl.pl. *puklā* ‘years’. However, *k* must be secondary in such examples because it effectively prevents the problematic vowel contractions in the morphologically expected forms $**lwes < *lwā.is$ (next to attested gen.sg. *lwes*!) and $**puklā < *puklā.ā$. Schmidt (cf. also Hartmann 2001) has argued that the *k* in roots ending in *-tk* goes back to $*h_2$, but Melchert's earlier derivation of *-tk-* from *-T-sk-* is definitely to be preferred (1977; cf. also Pinault 2006b). Kortlandt's derivation of Tocharian B *taka-* ‘be’ from $*steh_2-t$ with *-k-* from $*h_2$ is in itself attractive, but since the “*k*-aorist” is also attested in e.g. Gr. *ἔθηκα* and Lat. *fēcī*, this reflex cannot be used to determine the phylogenetic position of Tocharian, even if the

¹⁰ A thorough discussion of these developments can be found in Ringe 1996: 39–66 and Winter 1962. In both accounts, a complicating factor is the Tocharian version of Grassmann's Law, exemplified by e.g. Toch.B *tsik-* ‘form’ < $*d^heig^h$ and *tsāk-* ‘burn’ < $*d^heg^h$, allegedly with *ts* < $*d$ after $*d^h$ had been deaspirated to $*d$ before the following $*g^h$ and $*g^wh$, respectively. The evidence for Grassmann's Law in Tocharian is circumstantial and probably open to an alternative explanation. It is not taken into account here in view of the solid counterexample of Toch.B *tapre* ‘high’ < $*d^huh^hro-$, to be reconstructed with $*b^h$ instead of $*b$ after Kroonen (2011: 253, 255).

¹¹ It is possible that Tocharian inherited a stop system in which distinctive voice had not yet developed, as argued by Kortlandt (e.g. 1985: 197; 2020: 269), but in my view this is difficult to prove.

evidence as such nicely fits Kloekhorst's reconstruction of $*h_2$ and $*h_3$ as uvular stops for Proto-Indo-Anatolian (2018; cf. also Kortlandt 2002: 218).

Like other Indo-European languages, Tocharian shows reflexes of metathesis of $*Hi$ to $*iH$ and $*Hu$ to $*uH$. For instance, metathesis of $*Hu$ to $*uH$ is attested by such forms as Toch.B *puwar* 'fire' < $*puh_2r$ ¹² (as in Greek $\pi\acute{o}\rho$) from earlier $*peh_2ur$ (as in Hitt. *paḥḫur*) and Toch.B *ləw(a)-* 'rub' (prt.3sg.-3sg.obj. *lyawā-ne* 'he rubbed him') < $*leuh_3-$ from earlier $*leh_3u-$ (as in Hitt. *lāḫu-*ⁱ 'pour'). Even though unmetathesised forms are also found, for instance Toch.B *kaw-* 'kill' < $*keh_2u-$, the existence of metathesised forms in Tocharian clearly shows that this sound change is to be dated before Tocharian split off. However, even though Hittite often shows unmetathesised forms next to metathesised forms elsewhere (Kloekhorst & Pronk 2019: 5), the metathesis must have already occurred before Proto-Indo-Anatolian on the evidence of forms such as Hitt. *šuhḫa-* 'pour, sprinkle' < $*suh_2-$ next to *išḫu(wa)-* < $*seh_2u-$ and *lu-u-* 'pour' < $*luh_3-$ next to *lāḫu-* < $*leh_3u-$ (Melchert 2011: 129, 131). At this point, therefore, the mere attestation of laryngeal metathesis cannot be used for inner Indo-European phylogeny.

However, another Indo-European metathesis may be used: that of word-final $*-ur$ to $*-ru$ (Lubotsky 1994: 99–100). This sound change seems to have occurred only after Proto-Indo-Anatolian. Strong evidence for it in Tocharian has been discovered by Del Tomba (2021), who shows that Toch.B plurals in *-wa* to nouns in *-r*, such as *tarkār* 'cloud', pl. *tārkarwa*, presuppose metathesis of $*-ur$ to $*-ru$ in the singular, on which the plural *-r-wa* < $*-ru-h_2$ was built. Although this sound change may be used for the phylogeny of Indo-European, it clearly groups Tocharian together with the non-Anatolian languages.

6.5.3 Morphology

Morphology is the domain that is often ascribed the highest potential to yield evidence for the phylogenetic position of Tocharian. Indeed, morphology meets two essential needs: it is constantly in the process of change, and, at the same time, shifts in function, though commonplace, are subject to

¹² The Tocharian word for 'fire' is variously reconstructed. Hackstein, for instance, reconstructs $*ph_2uōr$ (2017: 1314). It is, however, questionable whether $*h_2$ would be lost in this context, and whether the reconstruction of a collective ending $*-ōr$ for this etymon is warranted. A derivation of Toch.B *puwar* from $*puh_2r$ is the most straightforward. Winter (1965: 192) reconstructs the Tocharian A equivalent *por* as $*paur$ from unmetathesised $*peh_2ur$. This is phonologically possible but most difficult morphologically, since it is not clear what the distribution of these variants in the Proto-Tocharian paradigm might have been. It is therefore preferable to assume a development $*wa > o$ similar to $*we > o$ in *koṃ*, obl.sg. of *ku* 'dog', < $*kwen$ and $*iye > e$ in *kareṃām* 'laughing' < $*keriyemane$ (Hilmarsson 1989: 135; Hackstein 2017: 1314; Peyrot 2012a: 210).

limitations. Unfortunately, Tocharian morphology is heavily reorganised and its prehistory is often very obscure. Even worse is the fact that the reconstruction of Proto-Indo-European in exactly the relevant points is difficult and often disputed.

Without a doubt, the most prominent argument for phylogeny based on morphology that has been advanced comes from the Tocharian *s*-preterite. In the active of the Tocharian *s*-preterite, an element *s* is only found in the 3sg.: 1sg. *prek-uwa* ‘asked’, 2sg. *prek-asta*, 3sg. *prek-sa*, 1pl. *prek-am*, 2pl. *prek-as**, 3pl. *prek-ar*. This is reminiscent of the Hittite *hi*-preterite, which likewise has *-š* only in the 3sg.: 3sg. *ākkiš* ‘died’, 3pl. *aker* (Pedersen 1941: 146). There are two schools of thought to explain this correspondence. The first, most prominently voiced by Jasanoff (e.g. 2003: 204–5),¹³ takes the restriction of the *-s-* as an archaism of Anatolian and Tocharian, while the rise of the classical *s*-aorist through generalisation of the *-s-* from the 3sg. throughout the whole paradigm is a common innovation of the other Indo-European branches. According to the second one, the *-š* in Hittite is secondary, probably somehow from the *s*-aorist, while in Tocharian the *s*-preterite forms without *-s-* lost it due to the effects of sound law and analogy (Ringe 1990; Kortlandt 1994; Peyrot 2013: 503–7). The matter cannot be treated here in detail. Suffice to say that the assumption of loss of *-s-* accounts best for the inflection of the Tocharian preterite and its patternings with the subjunctive. At any rate, this famous case very clearly shows how different views on the reconstruction of Proto-Indo-European logically lead to different evaluations of arguments for phylogeny.

Another phylogenetic argument is based on the middle endings in *-r* (e.g. Ringe, Warnow & Taylor 2002; Ringe 1991: 98–9). It is widely held that the shorter middle endings 3sg. **-to* and 3pl. **-nto* were secondary endings in Proto-Indo-European, while the corresponding primary endings were originally 3sg. **-to-r*, 3pl. **-nto-r*, which were later replaced by 3sg. **-to-i*, 3pl. **-nto-i*, marked with the productive primary marker **-i* as found in the active endings. This would not be a valid argument for Indo-Tocharian, since the *r*-endings are also found in Italo-Celtic and Phrygian, but it would group Tocharian with the older branches.

However, a number of problems with this argument need to be noted:

- It is questionable as to whether the contrast between Toch.B pres. 3sg. *-tär*, 3pl. *-ntär* and pret. 3sg. *-te*, 3pl. *-nte* continues an original primary–secondary contrast, because the Tocharian preterite active endings do not continue the secondary endings of Proto-Indo-European. In the copula 3sg. *ste*, 3pl. *skente*, the endings *-te*, *-nte* are even used as present endings. Hackstein (1995: 273–5) explains these forms as original resultatives, i.e. “is” < “has

¹³ Two more recent contributions are Melchert (2015) and Jasanoff (2019).

become”, and notes that presentic readings of preterites are found elsewhere. However, it remains problematic as to why no shade of the past meaning has been preserved in 3sg. *ste*, 3pl. *skente*, and why the corresponding suffixed forms, such as 3sg.-1sg.obj. *star-ñ*, have present endings. This distribution is difficult to explain from an original difference in tense.

- The reconstruction of the primary middle endings 3sg. **-to-r*, 3pl. **-nto-r* is problematic itself. Indeed, Lat. *-tur*, *-ntur* point to **-tor*, **-ntor*. However, as Weiss (2009: 413) notes, Osc. 3sg. *-ter*, 3pl. *-nter* point to **-tro*, **-ntro*, and Umb. primary *-ter*, *-nter* vs. secondary *-tur*, *-ntur* suggests Proto-Italic primary **-tro*, **-ntro* vs. secondary **-tor*, **-ntor*. Likewise, the Old Irish deponent endings 3sg. *-thir*, 3pl. *-tir* point to **-tr-*, **-ntr-*, probably **-tro*, **-ntro*. Finally, Toch. *-tär*, *-ntär* cannot be derived from **-tor*, **-ntor* directly (cf. also Pinault 2010b). Ringe (1996: 86) discusses the change of **-or* to Toch. **-är*, but the 3rd person middle endings are his only evidence, against counterexamples such as Toch.B *markwer* ‘milk’, with suffix *-wer* < **-uor* as in the verbal abstract, e.g. *šesüwer* ‘eating’. A further counterexample seems to be *yeter* ‘felloe’, which on the evidence of the unpalatalised *-t* must reflect **-tor*.¹⁴
- The assumed replacement of well-marked middle paradigms ending in *-r* with the active marker *-i* is difficult to understand. What would be the motivation to do so? If endings are clearly marked to be primary, there seems no need to replace them. The greatest difficulty here is not the addition of the primary active marker *-i* – such additions are indeed found frequently in e.g. the perfect endings, such as OCS *vědě*, Lat. *vīdī*, or Toch.A *kärse* ‘I knew’ < **kärsa-a-i* – but the fact that the transparent middle ending **-r* should have been deleted.

In view of these problems, it is tempting to follow Kortlandt’s reconstruction (1981) of the middle endings as **-to*, **-nto* only, without contrast between primary and secondary endings.¹⁵ Apparently such contrasts were created independently in the different branches. In any case, the problematic specifics of the reconstruction of the middle endings make them difficult to use for phylogeny.

Another argument advanced by Ringe, Warnow & Taylor (2002: 117) is the thematic optative in **-o-ih-*, attested in Indo-Iranian, Greek, Balto-Slavic and Germanic, but not in Tocharian. Indeed, this may be a later innovation within Indo-European not shared by Tocharian. In Tocharian, there is only one variant

¹⁴ A possible alternative reconstruction would be **-ewer* with contraction of **ewe* to *e*.

¹⁵ The evidence of Anatolian seems compatible with an original **-to*, **-nto* without contrast between primary and secondary endings: synchronically, they are attested in Hittite as pres.3sg. *-tta*, 3pl. *-anta*. However, a derivation from **-tor*, **-ntor* neatly explains the rise of the present particle *-ri* from resegmentation after the loss of *-r* after **-ó-* (Yoshida 1990). The introduction of the particle *-ti*, to mark the preterite endings, i.e. 3sg. *-ttati*, 3pl. *-antati*, would be motivated in both scenarios.

of the optative suffix, *-i-* (*i* with preceding palatalisation), to be derived from **-ih₁-*.¹⁶ However, “present optatives”, synchronically imperfects, are unattested in Tocharian A, and they must consequently have been regularised secondarily in Tocharian B (Peyrot 2012b). Therefore, it is difficult to prove that e.g. Toch.B *pari** ‘he took’ goes back directly to **b^her-ih₁-t* (for **b^her-o-ih₁-t* elsewhere). In any case, since the thematic optative is not attested in Italo-Celtic, it cannot be used to show that Tocharian split off before that branch.

It has been argued that the combination of the Tocharian present participle in *-mane* with both active and middle finite inflection is an archaism: the verbal adjective **-mh₁no-* would originally have been indifferent for voice, very much like the **-nt*-participle in Anatolian (Kloekhorst & Pronk 2019: 3), and became specialised only later, after Tocharian split off, as the middle counterpart of the active **-nt*-participle (Pinault 2012: 229; Peyrot 2017: 339–40). However, I now think that this argument has to be abandoned in the light of a study by Friis (2021), who shows that traces of specifically middle use are preserved, which suggests that active use of *-mane* in Tocharian is secondary.¹⁷

A case from word formation in the grammatical domain is the interrogative stem in **m-* found in Anatolian and Tocharian (Hackstein 2004: 280–3; Pinault 2010a: 359; Peyrot 2019a: 195–9). A weak point of this argument is that the innovation of the other Indo-European languages would consist only in *loss* of the *m*-interrogative, while a strong point is the central position of this stem, paired only with **k^wi-* (**k^we-*, **k^wo-*), in the linguistic system. Thus, while the identifiability of this feature is low, its salience is nevertheless high.

6.5.4 *Lexicon*

Lexical evidence has been variously evaluated. Important papers adducing lexical evidence in support of an early split off of Tocharian are Schmidt 1992 and Winter 1997. This evidence, and the method as a whole, was criticised by Hackstein (2005: 172) and Malzahn (2016) amongst others. For lexical arguments, a distinction should be made between lexical replacements and semantic change.

¹⁶ The full grade variant **-ye-* < **-ieh₁-* may have been ousted by the zero grade variant through paradigmatic levelling, but it is also possible that the zero grade variant was generalised from *s*-aorist optatives with **-ih₁-* throughout (if the synchronic optatives of root subjunctives of class 1, such as Toch.B *parši* ‘may he ask’, are to be derived from *s*-aorist optatives, i.e. in this case **prék-s-ih₁-t*).

¹⁷ Thus, even though I cannot agree with the arguments adduced by Fellner & Grestenberger (2018), I do now concur with their main claim.

Arguments based on lexical replacement are especially difficult because the identifiability requirement is not easily satisfied: it is hard to prove that two words did not carry the same or a similar meaning. An example of such an argument is Anatolian (i.e. Luvian) and Tocharian (i.e. Toch.A) **uel(H)-* ‘die’ vs. **mer-* elsewhere (Ringe, Warnow & Taylor 2002: 99).¹⁸ Although **mer-* indeed acquired the meaning ‘die’ from ‘disappear’ after Indo-Anatolian (Kloekhorst & Pronk 2019: 3), and thus became a new word for the meaning ‘die’, the Luvian and Tocharian A words cannot be shown to represent the original word for ‘die’, let alone that it was ousted by the new **mer-* (see Malzahn 2016: 285–6).

Another example is **h₁eg^{wh}-* ‘drink’, well attested in Tocharian and Anatolian, as against **peh₃-* elsewhere (Ringe, Warnow & Taylor 2002: 99). This may indeed be a case of lexical replacement, i.e. the meaning ‘drink’ came to be expressed by a different word. However, the details are complicated: Hitt. *pāšⁱ* ‘swallow’ shows that **peh₃-* needs to be reconstructed for Proto-Indo-Anatolian, with possibly only a slightly different meaning; and Lat. *ēbrius* ‘drunk’ and Gr. *νήφω* ‘be sober’ show that **h₁eg^{wh}-* was preserved after Tocharian split off, possibly with a shift to ‘be drunk’ (Peyrot 2019b). Thus, the argument for lexical replacement remains fragile, while the best phylogenetic evidence is formed by the possible semantic developments of ‘drink’ to ‘be drunk’ for **h₁eg^{wh}-*, and ‘swallow’ to ‘drink’ for **peh₃-*. The attestation of the meaning ‘be drunk’ in Latin is favourable for the Indo-Tocharian hypothesis, because it suggests that this semantic change occurred after Tocharian split off, but before Italo-Celtic split off.

As a lexical argument based on semantics, Ringe, Warnow & Taylor (2002: 99) adduce **meǵh₂*, of which the Anatolian (e.g. Hitt. *mekk-*, *mekki-*) and Tocharian (e.g. Toch.B *māka*) reflexes mean ‘much, many’, as against ‘great’ elsewhere. The distribution is especially neat in this case, since the etymon is also attested in Italo-Celtic (OIr. *maige*, Lat. *magnus*, etc.) and Germanic (Goth. *mikils*). Here, the main problem is the requirement of unidirectionality: the meanings are contingent and a change from ‘great’ to ‘much’ is by no means unlikely.

References

- Adams, Douglas Q. 1988. *Tocharian historical phonology and morphology*. New Haven, CT: American Oriental Society.
 Adams, Douglas Q. 2013. *A dictionary of Tocharian B*. Amsterdam: Rodopi.
 Anthony, David W. 2007. *The horse, the wheel, and language*. Princeton, NJ: Princeton University Press.

¹⁸ Their argument about **ai-* ‘give’ cannot be upheld with the reconstruction of Hitt. *pāi* ‘gives’ as **h₁p-oi-e* by Kloekhorst (2006).

- Anthony, David W. & Don Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1. 199–219.
- Burrow, Thomas. 1935. Tocharian elements in the Kharoṣṭhī documents from Chinese Turkestan. *Journal of the Royal Asiatic Society* 1935. 667–75.
- Debaine-Francfort, Corinne. 1988. Archéologie du Xinjiang des origines aux Han. Première partie. *Paléorient* 14. 5–29.
- Debaine-Francfort, Corinne. 1989. Archéologie du Xinjiang des origines aux Han. Deuxième partie. *Paléorient* 15. 183–213.
- Del Tomba, Alessandro. 2021. Metathesis of PIE **-ur* in Tocharian. *Münchener Studien zur Sprachwissenschaft* 74. 51–85.
- Dragoni, Federico, Niels Schoubben & Michaël Peyrot. 2020. The Formal Kharoṣṭhī script from the Northern Tarim Basin in Northwest China may write an Iranian language. *Acta Orientalia Hungarica* 73. 335–73.
- Fellner, Hannes & Laura Grestenberger. 2018. Die Reflexe der **-nt-* und **-mh₁no-* Partizipien im Hethitischen und Tocharischen. In Elisabeth Rieken (ed.), *100 Jahre Entzifferung des Hethitischen*, 63–82. Wiesbaden: Reichert.
- Friis, Louise S. 2021. Tocharian B agent nouns in *-ntsa* and their origin. *Indo-European Linguistics* 9. 1–25.
- Hackstein, Olav. 1995. *Untersuchungen zu den sigmatischen Präsensstambildungen des Tocharischen*. Göttingen: Vandenhoeck & Ruprecht.
- Hackstein, Olav. 2004. From discourse to syntax. The case of compound interrogatives in Indo-European and beyond. In Karlene Jones-Bley et al. (eds.), *Proceedings of the fifteenth annual UCLA Indo-European Conference*, 257–98. Washington, DC: Institute for the Study of Man.
- Hackstein, Olav. 2005. Archaismus oder historischer Sprachkontakt. Zur Frage west-indogermanisch-tocharischer Konvergenzen. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 169–84. Wiesbaden: Reichert.
- Hackstein, Olav. 2017. The phonology of Tocharian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1304–35. Berlin: De Gruyter Mouton.
- Hartmann, Markus. 2001. Wiederum zu den tocharischen Verben auf *-tk-*. *Münchener Studien zur Sprachwissenschaft* 61. 95–117.
- Hilmarsson, Jörundur. 1989. *The dual forms of nouns and pronouns in Tocharian*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Hilmarsson, Jörundur G. 1991. *The nasal prefixes in Tocharian. A study in word formation*. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Isebaert, Lambert. 1980. *De Indo-Iraanse bestanddelen in de Tocharische woordenschat*. PhD dissertation, University of Leuven.
- Jasanoff, Jay H. 1987. Some irregular imperatives in Tocharian. In Calvert Watkins (ed.), *Studies in memory of Warren Cowgill (1929–1985)*, 92–112. Berlin: De Gruyter.
- Jasanoff, Jay H. 2003. *Hittite and the Indo-European verb*. Oxford: Oxford University Press.
- Jasanoff, Jay H. 2017. The impact of Hittite and Tocharian: Rethinking Indo-European in the 20th century and beyond. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 1, 220–38. Berlin: De Gruyter Mouton.

- Jasanoff, Jay H. 2019. The sigmatic forms of the Hittite verb. *Indo-European Linguistics* 7. 13–71.
- Kloekhorst, Alwin. 2006. Hittite *pai-/pi-* ‘to give’. *Indogermanische Forschungen* 111. 110–19.
- Kloekhorst, Alwin. 2016. The Anatolian stop system and the Indo-Hittite hypothesis. *Indogermanische Forschungen* 121. 213–47.
- Kloekhorst, Alwin. 2018. Anatolian evidence suggests that the Indo-European laryngeals **h₂* and **h₃* were uvular stops. *Indo-European Linguistics* 6. 69–94.
- Kloekhorst, Alwin & Tijmen Pronk. 2019. Introduction: Reconstructing Proto-Indo-Anatolian and Proto-Indo-Uralic. In Alwin Kloekhorst & Tijmen Pronk (eds.), *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*, 1–14. Leiden: Brill.
- Kortlandt, Frederik. 1981. 1st sg. middle **-H₂*. *Indogermanische Forschungen* 86. 123–36.
- Kortlandt, Frederik. 1985. Proto-Indo-European glottalic stops: The comparative evidence. *Folia Linguistica* 6. 183–201.
- Kortlandt, Frederik. 1994. The fate of the sigmatic aorist in Tocharian. In Bernfried Schlerath (ed.), *Tocharisch*, 61–5. Reykjavík: Málvísindastofnun Háskóla Íslands.
- Kortlandt, Frederik. 2002. The Indo-Uralic verb. In Rogier Blokland & Cornelius Hasselblatt (eds.), *Finnic-Ugrians and Indo-Europeans: Linguistic and literary contacts*, 217–27. Maastricht: Shaker.
- Kortlandt, Frederik. 2018a. Proto-Indo-European glottalic stops: The evidence revisited. *Münchener Studien zur Sprachwissenschaft* 71. 147–53.
- Kortlandt, Frederik. 2018b. The Indo-European *k*-aorist. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 137–42. Ann Arbor, MI: Beech Stave.
- Kortlandt, Frederik. 2020. The disintegration of the Indo-European language family. In Harald Bichlmeier, Ondřej Šefčík & Roman Sukač (eds.), *Etymologus: Festschrift for Václav Blažek*, 267–76. Hamburg: Baar.
- Kroonen, Guus J. 2011. False exceptions to Winter’s Law: On the effects of Kluge’s Law on the Proto-Germanic consonantism. In Tijmen Pronk & Rick Derksen (eds.), *Accent matters: Papers on Balto-Slavic accentology*, 251–61. Amsterdam: Rodopi.
- Lane, George S. 1966. On the interrelationships of the Tocharian dialects. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 213–33. Berkeley: University of California Press.
- Lane, George S. 1970. Tocharian: Indo-European and non-Indo-European relationships. In George Cardona et al. (eds.), *Indo-European and Indo-Europeans*, 73–88. Philadelphia: University of Pennsylvania.
- Lubotsky, Alexander. 1994. Avestan *θβōrəstar-* and the Indo-European root *√turk-*. *Die Sprache* 36. 92–102.
- Mallory, James P. 2015. *The problem of Tocharian origins*. Philadelphia: University of Pennsylvania.
- Malzahn, Melanie. 2016. The second one to branch off? The Tocharian lexicon revisited. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 281–92. Wiesbaden: Reichert.

- Melchert, H. Craig. 1977. Tocharian verb stems in *-tk-*. *Zeitschrift für vergleichende Sprachforschung* 91. 93–130.
- Melchert, H. Craig. 2011. The PIE verb for ‘to pour’ and medial **h₃* in Anatolian. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 22nd annual UCLA Indo-European Conference*, 127–32. Bremen: Hempen.
- Melchert, H. Craig. 2015. The Tocharian *s*-preterite. In Melanie Malzahn et al. (eds.), *Tocharian texts in context*, 127–35. Bremen: Hempen.
- Pedersen, Holger. 1941. *Tocharisch vom Gesichtspunkt der indoeuropäischen Sprachvergleichung*. København: Munksgaard.
- Peyrot, Michaël. 2008. *Variation and change in Tocharian B*. Amsterdam & New York: Rodopi.
- Peyrot, Michaël. 2012a. The Tocharian A match of the Tocharian B obl.sg. *-ai*. *Tocharian and Indo-European Studies* 13. 181–220.
- Peyrot, Michaël. 2012b. Tocharian ‘eat’ and the strong imperfect in Tocharian A. In Olav Hackstein & Ronald I. Kim (eds.), *Linguistic developments along the Silk Road: Archaism and innovation in Tocharian*, 85–119. Vienna: Österreichische Akademie der Wissenschaften.
- Peyrot, Michaël. 2013. *The Tocharian subjunctive: A study in syntax and verbal stem formation*. Leiden: Brill.
- Peyrot, Michaël. 2017. On the part of speech and the syntax of the Tocharian present participle. In Claire Le Feuvre, Daniel Petit & Georges-Jean Pinault (eds.), *Verbal adjectives and participles in Indo-European languages*, 327–41. Bremen: Hempen.
- Peyrot, Michaël. 2018. Tocharian B *etswe* ‘mule’ and Eastern East Iranian. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 270–83. Ann Arbor, MI: Beech Stave.
- Peyrot, Michaël. 2019a. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-European Linguistics* 7. 72–121.
- Peyrot, Michaël. 2019b. Indo-Uralic, Indo-Anatolian, Indo-Tocharian. In Alwin Kloekhorst & Tijmen Pronk (eds.), *The precursors of Proto-Indo-European: The Indo-Anatolian and Indo-Uralic hypotheses*, 186–202. Leiden: Brill.
- Pinault, Georges-Jean. 2006a. Retour sur le numéral ‘un’ en tokharien. *Indogermanische Forschungen* 111. 71–97.
- Pinault, Georges-Jean. 2006b. Sur l’évolution phonétique *tsk* > *tk* en tokharien commun. *Münchener Studien zur Sprachwissenschaft* 62. 103–56.
- Pinault, Georges-Jean. 2010a. Le pronom d’ipséité en tokharien. In Injoo Choi-Jonin, Marc Duval & Olivier Soutet (eds.), *Typologie et comparatisme: Hommages offerts à Alain Lemaréchal*, 351–65. Leuven: Peeters.
- Pinault, Georges-Jean. 2010b. On the *r*-endings of the Tocharian middle. In Ronald I. Kim et al. (eds.), *Ex Anatolia lux: Anatolian and Indo-European studies in honor of H. Craig Melchert on the occasion of his sixty-fifth birthday*, 285–95. Ann Arbor, MI: Beech Stave.
- Pinault, Georges-Jean. 2012. Sound laws and the suffix of the PIE “middle” participle. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphonemics*, 227–51. Munich: LINCOM.
- Pinault, Georges-Jean. 2013. Anatolien et tokharien: Des langues décisives pour la reconstruction indo-européenne. *Histoire Épistémologie Langage* 35. 13–44.

- Ringe, Don. 1990. The Tocharian active *s*-preterite, a classical sigmatic aorist. *Münchener Studien zur Sprachwissenschaft* 51. 183–242.
- Ringe, Don. 1991. Evidence for the position of Tocharian in the Indo-European family? *Die Sprache* 34. 59–123.
- Ringe, Don. 1996. *On the chronology of sound changes in Tocharian*. Vol. 1. *From Proto-Indo-European to Proto-Tocharian*. New Haven, CT: American Oriental Society.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Schmidt, Klaus T. 1988. Stellungsbedingte Konsonantierung von a_2 im Tocharischen? In Alfred Bammesberger (ed.), *Die Laryngalthorie und die Rekonstruktion des indogermanischen Laut- und Formensystems*, 471–80. Heidelberg: Winter.
- Schmidt, Klaus T. 1992. Archaismen des Tocharischen und ihre Bedeutung für Fragen der Rekonstruktion und der Ausgliederung. In Robert S. P. Beekes, Alexander M. Lubotsky & Jos Weitenberg (eds.), *Rekonstruktion und relative Chronologie*, 101–14. Innsbruck: Institut für Sprachwissenschaft.
- Schmidt, Klaus T. 2018. *Nachgelassene Schriften*. Vol. 1. *Ein westtocharisches Ordinationsritual*; Vol. 2. *Eine dritte tocharische Sprache: Lolanisch*. Ed. by Stefan Zimmer. Bremen: Hempen.
- Schoubben, Niels. 2021. Accent sign matters. The Niya Prakrit grapheme <ḱ> and its connection to Bactrian <ḥk>. *Journal Asiatique* 309. 47–59.
- Shishlina, Natalia. 2008. *Reconstruction of the Bronze Age of the Caspian Steppes: Life styles and life ways of pastoral nomads*. Oxford: Archaeopress.
- Tremblay, Xavier. 2005. Irano-Tocharica et Tocharo-Iranica. *Bulletin of the School of Oriental and African Studies* 68. 421–49.
- Warries, Abel R. In press. Towards a new comparison of the pre-Proto-Tocharian and pre-Proto-Samoyed vowel systems. *Indo-European Linguistics*.
- Weiss, Michael. 2020. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor, MI: Beech Stave.
- Winter, Werner. 1962. Die Vertretung indogermanischer Dentale im Tocharischen. *Indogermanische Forschungen* 67. 16–35.
- Winter, Werner. 1965. Tocharian evidence for Proto-Indo-European laryngeals. In Werner Winter (ed.), *Evidence for laryngeals*, 190–211. The Hague: Mouton.
- Winter, Werner. 1997. Lexical archaisms in the Tocharian languages. In Hans H. Hock (ed.), *Historical, Indo-European, and lexicographical studies: A festschrift for Ladislav Zgusta on the occasion of his 70th birthday*, 183–93. Berlin: Mouton de Gruyter.
- Yoshida, Kazuhiko. 1990. *The Hittite mediopassive endings in -ri*. Berlin: de Gruyter.

7 Italo-Celtic

Michael Weiss

7.1 Introduction

Many scholars have noted similarities between Italic (Chapter 8) and Celtic (Chapter 9). Schleicher (1858) was the first to posit an Italo-Celtic node between Proto-Indo-European and Celtic and Italic.¹ But in the 1920s Carl Marstrander and Giacomo Devoto questioned the validity of this subgrouping.² Scholarly opinion has varied ever since. It would be fair to say that Italo-Celtic is more debatable than any other higher order subgrouping, certainly much more so than Balto-Slavic.

7.2 Evidence for the Italo-Celtic Subgroup

Many features once cited in favor of Italo-Celtic unity are now seen to be archaisms. For example, the medial *r*-endings (Lat. *sequitur* ~ OIr. *sechithir* ‘follows’) were in the nineteenth century only known from Italic and Celtic, but the appearance of these endings in Anatolian (Hitt. mid.3sg. *-ttari*), and Tocharian (Toch.B mid.3sg. *-tär/-trä*) completely changed this picture. It is true, however, that it is only in Italic and Celtic that *-r* becomes a marker of middle diathesis, and only Celtic and Latin have created a mid.1pl. **-mor*.³ In the other branches continuing **-r* the suffix is limited to the primary middle endings only: Hittite prim. *-ttari* : sec. *-ttati*; Toch.B prim. *-tär* : sec. *-te*.⁴ Another feature now known to be an archaism is the *t*-less 3rd singular medial endings: OIr. *berair* ‘is carried’, Umb. *ferar* subj.mid.3sg. These forms are

¹ But it is usually Lottner (1861) who is credited with first positing Italo-Celtic. In fact, Schleicher beat him to it by a few years. Schleicher mentioned the *r*-middle forms, the *ā*-subjunctive, and the *ī*-genitive as well as much other material that was just wrong. Lottner (1861) added the formation of the superlative.

² Devoto 1929; Marstrander 1929. Some key discussions of the issue of Italo-Celtic: Watkins 1966; Campanile 1968; Cowgill 1970; Jasanoff 1997; Schrijver 2016; Zair 2018; see also Kortlandt 1981, 2007.

³ Note, however, that in Old Irish for the 1st plural imperative of deponent verbs *r*-less forms occur in the glosses, e.g. *seichem* ‘sequamur’. See Thurneysen 1946: 37.

⁴ But note that the secondary middle endings were not completely eliminated. Lat. 2sg. *-re* continues < **-so* and Venetic continued *-to* as a pret.act.3sg. ending (*donasto* ‘gave’)

matched by Hitt. *-ari* (*ēšari* ‘sits’) and relics in Vedic (*áduha*[*t*] ‘gave milk’). Of course, archaisms like this do not provide positive evidence for subgrouping, but they aren’t completely uninteresting either. In the case of the primary marker **-r*, we may note that the nearest groups to the east, Proto-Germanic, Proto-Balto-Slavic, Albanian, and Greek have all taken part in the innovation of replacing primary middle *-r* with primary active *-i* (e.g. Goth. *haitada* ‘is called’ < **-otoj*, Arc. Gr. *-toi*). The fact that the two most westerly branches escaped this innovation may not be fortuitous.⁵

In the realm of phonology there are a small number of innovative features that have been proposed as shared Italo-Celtic developments, but these are all problematic.

Both Italic and Celtic agree in the development of **CRHC* to *CRāC*: Lat. *grānum* ‘a grain’ < **ǵr̥h_xnom* vs. Goth. *kaurn* < PGmc. **kurna*,⁶ OIr. *lām* ‘hand’ < **p̥h₂meh₂*, but this apparent isogloss is complicated by fact that both Italic and Celtic show other outcomes for this sequence. In Italic **CRHC* becomes *CaRaC* under the accent, e.g. *palma* ‘palm of the hand’ < **palama* < **p̥h₂meh₂* (see Höfler 2017). In Celtic the outcome *CRaC* is found in a number of examples, which cannot be easily explained as morphological neo-zero-grades, e.g. OIr. *flaith* ‘rule’, MW *gwlat* ‘country’ < **ul̥h_xti*.⁷ It is difficult therefore to believe that the resolution of **CRHC* sequences happened in Proto-Italo-Celtic. Note in particular the disagreement between MW *gwreid* ‘roots’ < **uradī* < **ur̥h_xdih₂* and the morphologically nearly identical Lat. *rādīx* ‘root’.

A famous isogloss that does seem to hold up better is the long-distance assimilation of **p ... k^w* > **k^w ... k^w* seen in Lat. *quīnque*, OIr. *cóic*, OW *pimp* ‘five’ < **k^wenk^we* < **pénk^we*.⁸

Latin *quercus* ‘oak’ < **k^werk^wu-* < **perk^wu-* (cf. Langobardic *fereha* ‘aesculus’, Goth. *fairguni* neut. ‘mountain’) seems to show that in Italic the assimilation **p ... k^w* > **k^w ... k^w* preceded the change of **k^wu* > **ku*. But the Celtic place-name *Hercynia* ‘oak forest’ < **perkunia* seems to show that in Celtic the **k^wu* to *ku* change preceded **p ... k^w* > **k^w ... k^w*. Since there was no **k^w* to trigger dissimilation **p* developed regularly to $\emptyset$. This relative chronology,

⁵ Proto-Balto-Slavic may have taken part in this innovation since the athematic (active) endings go back to *i*-diphthongs (OPr. *asmai* ‘I am’, *assei* ‘you are’), which may originate in the primary middle endings, though this is controversial. But note that Slavic has retained relic forms that could go back to **-or* in OCS *kažьdo* ‘everyone’ < **k^wos* + **ǵ^hido(r)* ‘is expected’ (Majer 2012: 230) and OCS *lubo* ‘or’ < **leub^ho(r)* ‘is wanted’ (Majer 2015). For Albanian see Schumacher 2016: 386. For the potential relevance of archaisms retained by adjacent languages see Watkins’s discussion (1966: 30).

⁶ OIr. *grán* and the other Celtic forms might be loanwords from Latin.

⁷ See Zair 2012: 69–89 for discussion.

⁸ The Sabellic form for ‘five’ was **pompe*, but strictly speaking it is not possible to determine whether this is from **k^wenk^we* or **penk^we*. Venetic also probably had this change, as it would have to if it is Italic, to judge from the Istrian ethnonym *Quarqueni* (Plin. 3. 130) ‘people of the oak forest’?

taken at face value, suggests that the Italic and Celtic long-distance assimilations were independent changes. If, however, the dissimilation of **k^wu-* to **ku-* occurred already in Proto-Indo-European, as is likely, then one might suppose that the labiovelar had been analogically restored from an oblique stem form **perk^weu-* in the dialects ancestral to Latin, in which case no inference about differing relative chronologies of the sound changes can be drawn.⁹

In my 2009 book, I entertained the possibility that Italic and Celtic shared the change of **ū* to **ī* before yod, sometimes called Thurneysen's Law. But Zair (2009) has shown that the Celtic facts are amenable to a different interpretation. The Old Irish word for 'smoke' *dé*, gen. *diad* must go back to an immediate preform **diots*, gen. *diotos* with a short *i* from earlier **d^huh₂iots*, **d^huh₂otos*. Zair explains this as **uh_xiV- > *uiV- > *iyV-*. Fortson (2017: 838) argues therefore that Thurneysen's Law is a different phenomenon. But the whole complex of facts deserves more discussion than we can give it here. I limit myself to two observations. First, the forms of the verb 'to be' with an *ī* reflecting **b^huh_x-je-* cannot be explained by an Italo-Celtic rule (Lat. *fiō*, Osc. *fiēt*, OIr. *biid*, but MW *byd* points to a short **i*) because these forms are also found in Germanic and Balto-Slavic (OE consuetudinal present *bið*, Lith. pret. 3ps. *bit(i)*, OCS conditional *bi*).¹⁰ Second, while Latin is uninformative about the vowel quantity in prevocalic position, the Sabellic cognates of *pīus* point unambiguously to a short *i* (Umb. *pehatu*, Pael. *pes* etc.).¹¹ This raises the possibility that the development in Italic, like Celtic, was by way of a short vowel.

In the realm of morphology we may note first the thematic genitive in **-ī̄*: Ogham Ir. MAQQI 'son', Gaul. SEGOMARI 'Segomarus', Lat. AISCOLAPI 'Aesculapius'.¹² Although the building blocks of the **-ī̄* genitive appear to be Proto-Indo-European (see Weiss 2020a: 204), the complete integration into the thematic nominal paradigm is uniquely Italic and Celtic. And yet this cannot have been a Proto-Italo-Celtic innovation. It is clear that the replacement of the inherited thematic gen.sg. **-os̄io* happened in the individual Celtic and Italic languages. VOL **-osio* is well represented in Satrican VALESIOSIO and in Faliscan *euotenosio*. Lepontic *-oiso* is a probably transformation of **-os̄io* under the influence of the pronominal gen.pl. **-oīsōm*. This means that Latin and Celtic in the historical period have independently replaced an inherited ending with the same piece of morphology. This could hardly be a contact phenomenon.¹³ Most scholars agree that the origin

⁹ The paradigm of the word for oak must have preserved its second syllable labiovelar in some forms. Cf. *Querquerni* the name of a Celtic tribe of Gallaecia 'people of the oak forest'.

¹⁰ See Hill 2012 for these forms. Hill does not discuss the Italic forms.

¹¹ The Oscan form *pihiūi* may be morphologically different (< **piijō-*).

¹² On the Messapic genitive in *-aihi*, which is not related, see Weiss 2020a: 221, 494; Matzinger 2019: 37.

¹³ The first instance of *-ī̄* in Latin is from the fifth/fourth century BCE Muracci di Crepadosso in Latium (MORAI ESOM 'I am of Morra.') The first secure Celtic example is from the second century BCE. It's highly unlikely that the *-ī̄* morpheme could have been transferred from Latin to

of the $\bar{i}$ -genitive is to be sought in the so-called $vrk\bar{i}h$ suffix $*-ih_2$, which makes substantives with genitival meaning from thematic nouns. The question then arises what function could the $vrk\bar{i}h$ suffix have acquired in Italic and Celtic that made it a favorable candidate for eventually replacing the inherited thematic gen.sg.? Answering this question is difficult because we have no attested textual evidence from Italic or Celtic showing both the inherited genitive and the $vrk\bar{i}h$ suffix. A necessary mid-stage for the transformation of the $vrk\bar{i}h$ suffix-forms, which are substantives in Indo-Iranian, into an adnominal case form would be their use as adjectives. This would be another instance of the so-called weak adjective phenomenon in which an original substantivized form becomes an adjective. Could the reinterpretation of the $vrk\bar{i}h$ suffix-forms as adjectives be the shared Italo-Celtic innovation that laid the groundwork for the eventual independent emergence of the $\bar{i}$ -genitive?

The $\bar{a}$ -subjunctive: OIr. *bera* ~ Lat. *ferat* ‘carry’. Both Italic and Old Irish display a morpheme $\bar{a}$ used to form the subjunctive.¹⁴ In Latin this makes the subjunctive to thematic present stems, but relic forms of Old Latin and Sabellic show derivation from the root (*advenas*, *atulas*, Umb. *neiřhabas*). This must represent an old pattern. In Old Irish the a -subjunctive is formed to weak presents and strong presents ending in *b*, *r*, *l*, *m*, and *n* plus *agaid*.¹⁵ Class S 3 (nasal infix presents to set root) affix the suffix to the root with no nasal infix (*benaïd* ~ *bia*). There are two schools of thought on the Italo-Celtic or Italic and Celtic a -subjunctive. One view, the traditional one, identifies the morphemes of the two language families. The other view, originating with Rix (1977) and significantly improved by McCone (1991), derives the Insular Celtic a -subjunctive from $*-ase-$, either the desiderative morpheme $*-h_1se-$ (Rix) or s -aorist subjunctive morpheme added to laryngeal final roots (McCone). The advantage of the McCone view is that it allows both Old Irish subjunctives to be derived from a single Proto-Indo-European category. But the disadvantage is that the starting point for the a -subjunctive on this hypothesis would be the s -aorist subjunctive built to set roots; such a category, which is very sparsely attested in other Indo-European languages, would have to have become very successful in the prehistory of Celtic.

The superlative formant $*-ismmo-$: OIr. *tressam* ‘strongest’ < $*treksisamos$, MW *hynaf* ‘oldest’ < $*senisamos$, Lat. *maximus* ‘greatest’ < $*magisVmos$,

Gaulish and then from Gaulish to the ancestor of the Insular Celtic languages, which were already on the British Isles by this time.

¹⁴ The oft-cited Tocharian class V $\bar{a}$ -subjunctive (Toch.A *wekas* ‘will disappear’, Toch.B *mārsam* ‘will forget’) does not belong with the Italic and Celtic forms. PIE $*\bar{a}$ becomes CToch. $*\bar{a}$ (Toch.A *a*, Toch.B *o*, e.g. Toch.A *pracar*, Toch.B *procer* ‘brother’ < $*b^hrātēr$ < $*b^hreh_2tēr$). See Jasanoff 1994: 206–7.

¹⁵ Strong presents ending in a velar and dental form the subjunctive with $-s-$.

Pre-Samnite *φολαισμος* ‘best’ (see Cowgill 1970). Even strong opponents of Italo-Celtic like Marstrander admit the strikingness of this agreement. Marstrander (1929: 246) wrote:

Une forme tout à fait identique comme irl. *nessam*, osque *nessimo*- doit provenir d’une même source primitive; on ne saurait guère admettre qu’elle se soit développée indépendamment dans les deux langues. Mais il n’en suit pas nécessairement qu’elle ait pris naissance à un époque d’unité italo-celtique.

[An absolutely identical form like OIr. *nessam*, Osc. *nessimo* must derive from the same original source; it would hardly be possible to accept that it had developed independently in the two languages. But it does not necessarily follow that it arose in an era of Italo-Celtic unity.]

Marstrander thought the proto-form of the superlative suffix was **-sṃmo-* and of “haute antiquité” [“remote antiquity”], hence a shared inheritance. But we know today, thanks to Warren Cowgill, that the proto-form was in fact **-isṃmo-* and it is certain that **-isṃmo-* replaces the earlier superlative formant **-isto-* continued by Greek, Indo-Iranian, and Germanic, which was inherited into Italic as traces like *iuxtā* ‘nearest’ and probably *ioviste* ‘youngest’ and *sōlistimus* ‘most favorable’ show.¹⁶ Furthermore **-isto-* could have been remade as **-isṃmo-* under the influence of the well-attested suffix superlative **-ṃmo-*, which is normally added to pronominal and adverbial stems. But on what basis could a theoretical archaism **-isṃmo-* be remade to **-isto-*, since the suffix *-to-* would not otherwise occur as a superlative formant? The superlative formant **-isṃmo-* seems the strongest argument for Italo-Celtic. It should be noted, by the way, that the same formant is continued in (para-) Venetic (VENIXEMA from Emona), but this is unproblematic if one believes, as I do, that Venetic was an Italic language.

Primary 3rd person middle endings **-tro*, **-ntro*: OIr. *do.moinethar* ‘thinks’, Umb. *herter* ‘should’ < **her(i)tro*.¹⁷ The ending **-ntro* results from a contamination of **-ntor* and **-ro* and the innovation spread from the 3rd plural to the 3rd singular. This innovation did not succeed in completely ousting

¹⁶ For possible traces of the superlative suffix **-isto-* in Celtic personal names, see Prosper 2018: 128–9. Some reconstruct a laryngeal after the **t* because of Ved. *-iṣṭha-*.

¹⁷ The source for the Old Irish deponent 3rd singular and plural endings and the Umbrian primary middle endings must be reconstructed as **-trV*, **-nttrV*. If the ending had been **-tor*, a pre-OIr. **sekʷitor* ‘follows’ would have syncope the medial vowel. The attested form *sechithir* points to an immediate preform **sekʷitr*. See Thurneysen 1946: 367 and Jasanoff 1997. Final **(n)tro* in Umbrian and Oscan became [t̪r]. In Oscan the new vowel merged with the reflex of short *i* and is consistently written with *e*. In Umbrian the vowel merged with the reflex of short *i* and is written with *e* in the Umbrian alphabet and *e*, *i*, or *ei* in the Latin alphabet. Meiser (1986: 112) champions Ebel’s suggestion to derive the forms from **-ti-r* and **-ntir* with an *r* tacked on to the primary active personal endings, but this is unnecessary because there is just not enough evidence to show that the outcome of final **Cros* was anything different. On Umb. *ocar*, which is from **okaris* not **okris*, see Weiss 2013: 349.

*-tor and *-ntor in either Italic or Sabellic. In any case, there is no evidence for this contamination elsewhere in Indo-European.

At a much lower level of importance are the many shared lexical items, since content words can be easily borrowed. Nevertheless, some of these items show striking morphological and semantic specializations. Some examples follow.

Lat. *crispus* ‘curly’, MW *crych*, Gallo-Lat. PN *Crixsus* continue a proto-form **kripso-* from the root **kreip-* ‘turn’ found also in Balto-Slavic (OCS *krěsъ* ‘solstice’, Lith. *kreipti* ‘to turn’). The Italic, Celtic, and Slavic forms presuppose an *s*-stem **kreipos* ‘turning’. In Proto-Italo-Celtic the *s*-stem made a thematic derivative, which, in the most archaic fashion, triggered a double zero-grade of the pre-suffixal stem. The meaning ‘having turning’ was specialized to ‘curly’ and ‘wrinkled’, both meanings attested in Welsh and Latin.¹⁸

Lat. *dēses*, *dēsidis* ‘lazy’, ‘inactive’, OIr. *deeid* < **de-sed(i)-*. The Latin adjective, which is not attested before Livy, has been suspected of being backformed from *dēsidia* ‘idleness’ (Plautus +), but the close match with the Irish adjective makes this unlikely. The Irish and the Latin form presuppose a semantic development **de/deh₁* + **sed-* ‘to remain seated’ (cf. Lat. *dēsideō*) > ‘to be idle’.

Lat. *saeculum* ‘lifespan’, MW *hoedl* ‘lifetime’ < **saīt₂lom* < **seh₂itlom*, Gaul. *deae setloceniae* < **saītlokejnijo-* ‘goddess of long life’ (cf. OIr. *cían* ‘long’). This match is perfect and, if correctly derived from the root **seh₂i-* ‘bind’, shows a striking semantic development. The oldest recoverable meaning for both *hoedl* and *saeculum* is ‘lifespan’. Thus in early Rome, according to Etruscan belief, a *saeculum* extended from some important date like the founding of Rome until the last person alive at that initial time died. This meaning could have arisen from the idea of a binding knot, marking the ends of life. Cf. Ved. *pārur-* ~ *pārvan-* which means ‘a knot’, ‘a limit’ and also ‘a fixed period of time’.

Lat. *dē* ‘down from’, OIr. *di*, OW *di*. This preposition, probably the instrumental **deh₁* of a pronominal stem **do-*, has no precise matches outside of Italic and Celtic. Though just a little word, **deh₁*’s import is considerable since it is part of a relatively small set of quasi-functional prepositions.

¹⁸ De Vaan (2008: 145) prefers a proto-form **krispo-* which is equally possible on the grounds that **kris-* is attested in Latin in *crinis* ‘hair’ and *crista* ‘crest’, but neither *crispus* nor *crych* is exclusively a descriptor of hair, and it is easier to explain an *-s-* as a remnant of an old *s*-stem than a *-p-* as a root extension.

Lat. *Sēmō*, a god of the oath often associated with Hercules, and Osc. **seemún-** match Gaul. *Segomon-*, an epithet of Mars. These forms converge on a Proto-Italic epithet **seǵʰo-mō*, *-mon-* ‘strong-man’, a secondary *-mon-*stem from a thematic stem **seǵʰo-* ‘strength’ (Mlr. *segh*). The form **seǵʰo-mō* seems to have been a divine epithet found nowhere but in Italic and Celtic (see Weiss 2017a).

Whether one recognizes an Italo-Celtic node or not, the fact remains that Italic shares more innovative features with Celtic than with any other branch.¹⁹ Nevertheless, it should not be forgotten that both Italic and Celtic individually and in common share many features with Germanic. This connection is not surprising given their geographical positions (see Weiss 2020a: 500–1). Somewhat more surprising are some striking agreements between Italic and/or Celtic and Indo-Iranian, famously highlighted by Vendryes. The phylogenetic import of these agreements is still unclear (see Weiss 2020b).

7.3 The Position of Italo-Celtic²⁰

The relationship of Italo-Celtic to the rest of Indo-European can be conceived of as the answer to three questions. (1) Was Proto-Italo-Celtic the next clade to separate from the PIE tree after the separation of Proto-Tocharian? (2) How do we interpret the extensive lexical matches between Italic, Celtic, and the other northern Indo-European branches, Germanic and Balto-Slavic, the so-called vocabulary of the northwest? (3) What do we make of the striking matches,

¹⁹ For a determined attempt to undermine the plausibility of Proto-Italo-Celtic from the phonological side, see Isaac 2007: 75–95. His argumentation is based on very specific possible formulations of the sound changes and, consequently, relative chronologies which, in my opinion, either can be formulated differently or cannot be stated with sufficient certainty. For example, Isaac relies heavily on the failure of the word for ‘yesterday’ to fall together with the reflect of “thorn” clusters in Italic (**gʰdʰ(i)es-* > Lat. *heri*). If the metathesis of *TK* to *KT* is Proto-Italic-Celtic (or earlier) **gʰdʰies* would have to become **dʰgʰ(i)es*. But if this is the case, then how did the Latin form escape the normal treatment of such clusters in Latin to initial *s-* (*situs* ‘decay’ < **dʰgʰiitu-*). One solution, Isaac suggests, is to posit a simplification of **gʰdʰ(i)es* to **gʰ(i)es* in Proto-Italic but not in Proto-Celtic where the outcomes with *d* (OIr. *indé*, W *doe*) show that this simplification could not have applied. This difference would necessarily mean that Proto-Italic was divergent from Proto-Celtic at this point and the metathesis, if shared by Italic and Celtic, would be a diffused or independent event. By Isaac’s chronology there would then be no unique phonological innovations shared between Italic and Celtic predating this divergence. But this assumes that the thorn cluster development was the result of simple metathesis. In fact, what if, as argued by Jasanoff (2018), the key to the thorn cluster development was spontaneous palatalization in *TK* clusters with subsequent metathesis? i.e. *TK* > *TʰKʰ* > *KT*. If this was the development, then there is no necessity for *KT* to have the same development as *KTʰ*. In some languages these might have merged and in others, including Latin, they did not.

²⁰ For the sake of this exposition, I will take the validity of the Proto-Italo-Celtic subgroup for granted.

especially in the religious and legal lexicon, shared by Italo-Celtic and Indo-Iranian?

That Proto-Italo-Celtic was the next group to branch off after Proto-Tocharian has been supported by some computational phylogenies of Indo-European (see Figure 7.1) but not others.²¹ To show that Proto-Italo-Celtic was the next to branch off would require demonstrating the existence of innovations shared by all the other non-Anatolian, non-Tocharian branches that are not found in Proto-Italo-Celtic. The best candidate for an innovation of this sort is the thematic optative **-o-ih₁-* of which there is no certain trace in Italic or Celtic, while it is well represented, or at least traceable, in Germanic, Balto-Slavic, Indo-Iranian, Greek, Armenian, Phrygian, and Messapic.²² In place of the thematic optative, on the view followed here, Italic and Celtic show the **ā*-subjunctive. Another possible innovation of the inner branches is the replacement of the primary middle marker **-r* by **-i*, which is seen in Greek, Phrygian,²³ Indo-Iranian, Germanic, Albanian, and possibly Balto-Slavic.

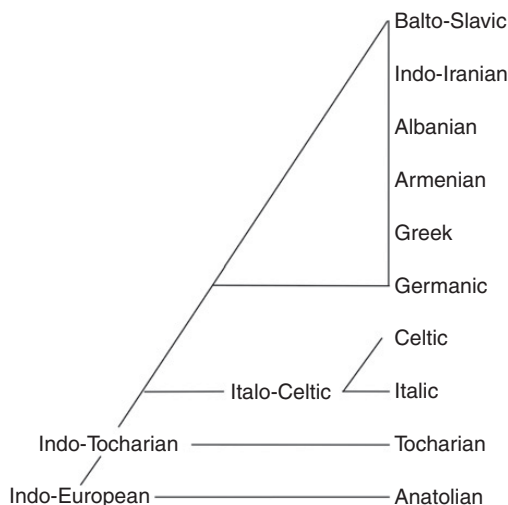


Figure 7.1 Tentative tree showing the position of Italo-Celtic

²¹ This is the finding of Ringe, Warnow & Taylor 2002, but it is not supported by the Chang et al. 2015 tree.

²² There is no indisputable evidence for the retention of the thematic optative in Albanian, but, given its advanced state of development at time of first attestation, this is not too surprising.

²³ Old Phrygian has only *-toi*. New Phrygian has two instances of a 3sg. sequence *-tor*. It's not clear that these are to be compared with the *r*-middle forms of Anatolian, Tocharian, and Italo-Celtic.

These two potential isoglosses seem to constitute the total evidence for innovations not reaching Proto-Italo-Celtic.

At the same time, it is clear that Proto-Italo-Celtic was in close contact with the rest of the northwestern Indo-European branches. Meillet (1922) famously identified a long series of lexical items shared between Italic, Celtic, Germanic, Baltic, and Slavic that found no matches in the other IE languages (cf. also Oettinger 2003). With greater knowledge of Anatolian, Tocharian, and the later Iranian languages, some of these supposedly exclusive items must be reevaluated. For example, the root **seh₁-* ‘sow’ (Lat. *sēmen* ‘seed’, OIr. *síl*, OHG *sāmo*, OCS *sěmę* ‘seed’, Lith. *sėti* ‘to sow’) now has a cognate in Hitt. *šāi*, *šīyanzi* ‘to press’. The item **seh₁-* must be reconstructed for highest node PIE, but the specialization to ‘sow’ is still only found in the northwest. On the other hand, Meillet’s example **porkōs* ‘piglet’ (Lat. *porcus*, OHG *farah*, Lith. *pařšas*, CS *prasę*) is no longer valid since a cognate is attested in Iranian (YAv. *parsa-*, Khot. *pāsa*, etc.).

Nevertheless, there are still many items with a northwestern distribution. Some of these might be common or independent borrowings from substratal languages. This scenario is especially plausible for the names of flora and fauna. An example of this sort might be ‘alder’. The cognates for this word show a remarkable amount of formal variation that is difficult to trace back to exclusively Indo-European morphophonology: Lat. *alnus* < **alsno-*; PGmc. **alisō* (ODu. *elis* in place-names; MDu. *else*, Sp. *aliso*) ~ **alizō* (OHG *elira*) ~ **aluz-* (ON *qlr*, OE *alor*); Lith. *aliksiņis* *ālksnis*, *elksnis*; PSI. **olbxa* (Ru. *ol’xá*) ~ **elbxa* (Ru. dial. *elxá*, Bulg. *elxá*) ~ **olbša* (Cz. *olše*) ~ **eliša* (SCR. *jěľša*). Cf. Basque *haltz*. The word may, however, also show up in Macedonian *ăľičă* (Hsch.) glossed as ‘poplar’.

Two terms relating to agricultural technology with somewhat overlapping meanings are (1) **l(V)ih_xseh₂* ‘furrow, track’ (Lat. *līra* ‘furrow’ < **lei_xseh₂*; OPr. *lyso* ‘field’ < **lih_xseh₂*, cf. Lith. *lysė* < **lih_xsi_jeh₂*; OCS *lěxa* ‘row’, OHG *-leisa* ‘track’ < **loi_xseh₂*²⁴) and (2) **polkeh₂* ‘ploughed piece of land’ (OE *fealh* ‘ploughed land’, Gaul. **olca* ‘arable land’ (Gregory of Tours *olca*, OFr. *ouche*, Port. *olga*), ORu. *polosá* ‘strip of land’). In Latin, Germanic, and Slavic the root **plek-* ‘plait’ has acquired a *-t*-extension: Lat. *plectere*, OHG *flehtan*, OCS *pletę*. Contrast the unextended **plek-* in Lat. *ex-plicere* and Gk. *πλέκω*. A piece of military technology is reflected by the word for ‘shield’: Lat. *scūtum*, OPr. *staytan* for **skaitan* < **sko_jitom* vs. OIr. *sciath*, MW *ysgwyd*, OCS *štiti* < **ske_jitom*.

There are a number of words relating to social structure. Most famous is the word **teuteh₂* ‘people’ (Osc. *touta*, Goth. *þiuda*, OIr. *tuath*, Lith. *tautà*). And in

²⁴ Whether these forms are further connected with the root **leis-* ‘learn’ (LIV² 409) is doubtful, but in any case, the agricultural meaning is a share feature of the northwest.

quasi-opposition to **teuteh₂* is **g^hostis* ‘guest-friend’ (Lat. *hostis*, Ven. *hosti-hauos*, Goth. *gasts*, OCS *gostъ*). From the legal sphere we have **d^helg^h-* ‘owe’ (OIr. *dligid* ‘is owed’, OIr. *dliged* ‘law’, Goth. *dulgs* ‘debt’, OCS *dlъgъ* ‘debt’, though the Slavic forms might be a loan from Gothic) and **uad^h-* ‘surety’ (Lat. *vas*, *vadis*, Osc. **vaamunim** ‘vadimonium’ < **uafemōnīom*, Goth. *wadi* ‘pledge, surety’, Lith. *vādas* ‘surety’ (obsolete)).²⁵

Finally, it’s been observed since Vendryes 1918 that Italo-Celtic and Indo-Iranian share a number of culturally important words relating to the religio-legal sphere not occurring in the intervening languages. The most notable of these are the words **h₃rēgs* ‘rule’, ‘king’ (OIr. *rí*, Lat. *rēx*, Ved. *rāt*) and **kred(s)-d^heh₁-* ‘to trust’, lit. ‘place heart’ (OIr. *creitid*, Lat. *crēdere*, Ved. *śraddhā* ‘trust’). Vendryes regarded these agreements as archaisms that were discarded in the intermediate languages, but it is striking that the supposed archaic status of these items is not confirmed by evidence from Proto-Anatolian or Proto-Tocharian.²⁶

References

- Campanile, Enrico. 1968. Studi sulla posizione dialettale del latino. *Studi e saggi linguistici* 8. 16–130.
- Chang, Will, Chundra Cathcart, David Hall, & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91. 194–244.
- Cowgill, Warren. 1970. Italic and Celtic superlatives and the dialects of Indo-European. In George Cardona, Henry M. Hoenigswald, & Alfred Senn (eds.), *Indo-European and Indo-Europeans*, 113–53. Philadelphia: University of Pennsylvania Press.
- Devoto, Giacomo. 1929. Italo-greco e italo-celtico. In *Silloge linguistica dedicata alla memoria di Graziadio Isaia Ascoli nel primo centenario della nascita*, 200–40. Turin: Chiantore.
- Fortson, Benjamin W., IV. 2017. The dialectology of Italic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of historical and comparative Indo-European linguistics*. Vol. 2, 835–57. Berlin: De Gruyter Mouton.
- Hill, Eugen. 2012. Die Entwicklung von **u* vor unsilbischem **i* in den indogermatischen Sprachen Nord- und Mitteleuropas: Die Stammsuppletion bei *u*-Adjektiven und das Präsens von ‘sein’. *NOWELE* 64/65: 5–36.
- Höfler, Stefan. 2017. Observations on the *palma* rule. *Pallas: Revue d’études antiques* 103. 15–23.

²⁵ Some regard **uad^h-* as a form of the root **ued^h-* ‘lead’, but the semantic and formal differences argue for keeping them apart.

²⁶ On this issue, see Weiss 2017b and 2020b. A word must also be said about Lusitanian, which is known only from a small number of inscriptions from the Western Iberian Peninsula. Prósper (2021) has made a convincing case that Lusitanian is not Celtic but if anything more closely related to Italic. We eagerly await the discovery of more inscriptions that might answer the question of the affinities of Italic and Lusitanian more conclusively.

- Isaac, Graham. 2007. *Studies in Celtic sound changes and their chronology*. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Jasanoff, Jay H. 1994. The Brittonic subjunctive and future. In Jens Elmegård Rasmussen (ed.), *In honorem Holger Pedersen*, 199–220. Wiesbaden: Reichert.
- Jasanoff, Jay H. 1997. An Italo-Celtic isogloss: The 3pl. mediopassive in **-ntro*. In Douglas Q. Adams (ed.), *Festschrift for Eric Hamp*. Vol. 1, 146–61. Washington, DC: Institute for the Study of Man.
- Jasanoff, Jay H. 2018. Palatable thorns. In Dieter Gunkel et al. (eds.), *Vina diem celebrant: Studies in linguistics and philology in honor of Brent Vine*, 133–40. Ann Arbor, MI: Beech Stave.
- Kortlandt, Frederik. 1981. More evidence for Italo-Celtic. *Ériu* 32. 1–22.
- Kortlandt, Frederik. 2007. Italo-Celtic. In *Italo-Celtic origins and prehistoric development of the Irish language*, 149–57. Amsterdam: Rodopi.
- Lottner, C. 1861. Celtisch-italisch. *Beiträge zur vergleichenden Sprachforschung auf dem Gebiete der Arischen, Celtischen und Slawischen Sprachen* 2. 309–21.
- Majer, Marek. 2012. An archaic Indo-European verbal form in the Slavic generalizing particle **-žido*? In H. Craig Melchert (ed.), *The Indo-European verb*, 225–34. Wiesbaden: Reichert.
- Majer, Marek. 2015. A lovely alternative: Proto-Slavic **ljubo*. In E. Mańczak-Wohlfeld & B. Podolak (eds.), *Words and dictionaries: A festschrift for Professor Stanisław Stachowski on the occasion of his 85th birthday*, 193–203. Cracow: Jagiellonian University Press.
- Marstrand, Carl. 1929. De l'unité italo-celtique. *Norsk Tidsskrift for Sprogvidenskap* 3. 241–68.
- Matzinger, Joachim. 2019. *Messapisch*. Wiesbaden: Reichert.
- McCone, Kim. 1991. *The Indo-European origins of the Old Irish nasal presents, subjunctives and futures*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Meillet, Antoine. 1922. *Les dialectes indo-européens*. 2nd ed. Paris: Champion.
- Meiser, Gerhard. 1986. *Lautgeschichte der umbrischen Sprache*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Oettinger, Norbert. 2003. Neuerungen in Lexikon und Wortbildung des Nordwest-Indogermanischen. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in prehistoric Europe*, 183–93. Heidelberg: Winter.
- Prósper, Blanca María. 2018. The Indo-European personal names of Pannonia, Noricum, and Northern Italy: Comparative and superlative forms in Celtic, Venetic, and South-Picene. *Voprosy onomastiki* 15. 108–38.
- Prospér, Blanca María. 2021. The Lusitanian oblique cases revisited: new light on the dative endings. In Juan Antonio González Iglesias, Julián V. Méndez Dosuna, & Blanca María Prósper (eds.), *Curiositas nihil recusat*: Studia Isabel Moreno Ferrero dicata, 427–42. Salamanca: Universidad de Salamanca.
- Ringe, Don, Tandy Warnow, & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Rix, Helmut. 1977. Das keltische Verbalsystem auf dem Hintergrund des indo-iranisch-griechischen Rekonstruktionsmodells. In Karl Horst Schmidt (ed.), *Indogermanisch und Keltisch*. 132–58. Wiesbaden: Reichert.

- Schleicher, August. 1858. Die stellung des celtischen im indogermanischen sprachstamme. *Beiträge zur vergleichenden Sprachforschung auf dem Gebiete der arischen, celtischen und slawischen Sprachen* 1. 437–48.
- Schrijver, Peter. 2016. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic. In John T. Koch & Barry Cunliffe (eds.), *Celtic from the West*. Vol. 3, 489–502. Oxford: Oxbow.
- Schumacher, Stefan. 2016. The development of the PIE middle in Albanian. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 385–400. Wiesbaden: Reichert.
- Thurneysen, Rudolf. 1946. *A grammar of Old Irish*. Revised edition with supplement by D. A. Binchy and O. Bergin. Dublin: Dublin Institute for Advanced Studies.
- de Vaan, Michiel. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.
- Vendryes, Joseph. 1918. Les correspondances de vocabulaire entre l'indo-iranien et l'italo-celtique. *Mémoires de la société de linguistique* 20. 265–85.
- Watkins, Calvert. 1966. Italo-Celtic revisited. In Henrik Birnbaum & Jaan Puhvel (eds.) *Ancient Indo-European dialects*, 29–50. Berkeley: University of California Press.
- Weiss, Michael. 2009. *Outline of the historical and comparative grammar of Latin*. Ann Arbor, MI: Beech Stave.
- Weiss, Michael. 2013. Interesting *i*-stems in Irish. In Adam I. Cooper et al. (eds.), *Multi nominis grammaticus*, 329–45. Ann Arbor, MI: Beech Stave.
- Weiss, Michael. 2017a. An Italo-Celtic divinity and a common Sabellic sound change. *Classical Antiquity* 36. 370–89.
- Weiss, Michael. 2017b. King: Some observations on and east-west archaism. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Usque ad radices: Indo-European studies in honour of Birgit Anette Olsen*, 793–800. Copenhagen: Museum Tusculanum.
- Weiss, Michael. 2020a. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor: Beech Stave.
- Weiss, Michael. 2020b. The inner revolution: Old but not that old. In Matilde Serangeli & Thomas Olander (eds.), *Dispersals and diversification: Linguistic and archaeological perspectives on the early stages of Indo-European*, 263–86. Leiden: Brill.
- Zair, Nicholas. 2009. OIr. *biid* < **bhuH*-ye/o- and “hiatus” verbs. In Stephanie Jamison, H. Craig Melchert, & Brent Vine (eds.), *Proceedings of the 20th UCLA Indo-European Conference*, 213–20. Bremen: Hempen.
- Zair, Nicholas. 2012. *The reflexes of the Proto-Indo-European laryngeals in Celtic*. Leiden: Brill.
- Zair, Nicholas. 2018. The shared features of Italic and Celtic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 2030–7. Berlin: De Gruyter Mouton.

8 Italic

Michael Weiss

8.1 Introduction

The Italian peninsula before the Roman conquest was home to a large number of languages, both Indo-European and non-Indo-European.¹ Among these languages, the following have been thought to descend from a common ancestor, Proto-Italic (cf. [Figure 8.1](#)).

1. Latin, spoken in Latium in a number of slightly divergent dialects for most of which we have only scant information from inscriptions and glosses. The Latin of Praeneste, which is the findspot for the two earliest Latin inscriptions, and the Latin of Falerii are reasonably well attested in inscriptions dating from the seventh to second century BCE. The Latin of Falerii is often classified as an independent language called Faliscan, though this is not justified on linguistic grounds.² But towering above all is the Latin of Rome. In this language we have a small number of inscriptions from the seventh–sixth centuries BCE in a distinctively archaic form, which I call Very Old Latin. After slowing to a trickle in the fifth century, Latin inscriptions pick up again in the fourth century and are joined by literary documents in the third century. The Latin of Rome spread first to Italy, suppressing the previously existing linguistic diversity, and then to most of Western Europe, North Africa, and southeastern Europe north of the Jireček line.³ Roman Latin survives today in its multiple descendants, the Romance languages.
2. The Sabellic languages. These languages, which form an as yet unquestioned subgroup, are
 - a. Oscan, the language of the Samnites of central and southern Italy, who also expanded into Campania and Sicily, is represented by about 800 inscriptions dating from the mid-fourth century BCE to perhaps as late as the first century CE

¹ See [Weiss 2020](#): 15–18 for a survey.

² On Faliscan see [Bakkum 2009](#). There are about 360 linguistically informative Faliscan inscriptions dating from the sixth to second centuries BCE. On Praenestine, see [Franchi de Bellis 2005](#).

³ An imaginary line drawn by the historian Konstantin Jireček marking the southern extent of Latin influence in southeast Europe.

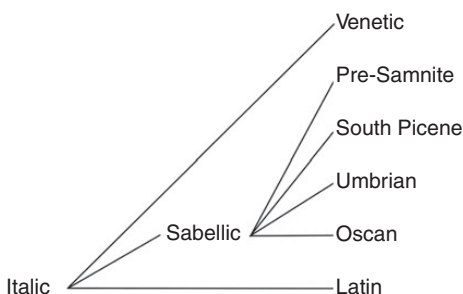


Figure 8.1 The Italic languages

- b. Umbrian, known chiefly from the Iguvine Tables from Gubbio (third–second century BCE; see [Weiss 2010](#)) and about forty smaller inscriptions, a few as early as the seventh century
- c. South Picene,⁴ the language of fewer than thirty inscriptions from the Marche and Abruzzo dating from the sixth–third centuries BCE
- d. Pre-Samnite, the language of inscriptions from Campania before the Samnite conquest in the fifth century; the longest document is the Cippus of Tortora from the sixth–fifth centuries⁵
- e. In addition, there are a number of short texts in the dialects of the Volsci, Marsi, Paeligni, Marrucini, Vestini, and Hernici.⁶ We also have a number of Sabellic loanwords in Latin (*bōs* ‘cow’ for expected **(w)ūs* < **gʷous* being the most prominent of them).⁷
3. Venetic, attested in more than 400 inscriptions from the northeast corner of Italy from the sixth to first centuries BCE. Some documents have been discovered in neighboring Slovenia and Austria. Not all scholars would agree that Venetic is an Italic language.
4. Sicel, the language of a small number (fewer than thirty) of pre-Greek inscriptions of eastern Sicily from the sixth to fourth centuries BCE and a number of glosses.⁸ It is very difficult to determine much about this language beyond that it was Indo-European, as the form *pibe* ‘drink!’ = Ved. *piba* shows

⁴ See [Zamponi 2021](#) for a survey of the evidence.

⁵ But see [Clackson 2015](#): 26–7 who questions Rix’s idea of uniting these texts and some inscriptions of Lucania as a unitary language. [Adiego \(2015\)](#) prefers Opic for the language of these inscriptions, which must have coexisted with Oscan for some time after the Samnite invasion of Campania.

⁶ There is no space to discuss the internal subgrouping of Sabellic, which is also not uncontroversial. See [Clackson 2015](#) and [Fortson 2017](#): 847–51.

⁷ See [Pocchetti 2017](#) for more detail.

⁸ On Sicel, see [Ambrosini 1984](#), [Agostiniani 1992](#), [Campanile 1969](#), [Willi 2008](#): 341–8, [Pocchetti 2012](#), and [Hartmann 2018](#).

conclusively.⁹ There are a few items in Sicilian Doric Greek that seem to match Latin and that are suspected of being of Sicel origin, e.g. *litra* ~ Lat. *libra* ‘pound’, *kūbiton* ~ Lat. *cubitum* ‘elbow’. The few inscriptions that are longer may show some Italic lexical material such as Mendolito *geped* ‘had’,¹⁰ with a simple perfect comparable to Osc. *hipid*, Grammichele *dedaxed* ‘made’ (?) (see Machajdiková 2018: 151), perhaps with a reduplicated *k*-extended form of the root **d^heh₁-* like VOL *fhe:fhaked*, Osc. *fefacid*, or the female name *Kup(a)ra*, which recalls Sabellic **kupro-* ‘good’. If Sicel is Italic, it would diverge from all other members in showing voiced reflexes of the PIE voiced aspirates in initial position in contrast to the *f* < **d^h*, **b^h* and *h* < **g^h*/**ǵ^h* seen in Latin, Sabellic, and Venetic.

8.2 Evidence for the Italic Branch

Positing Proto-Italic as the superordinate node of Latin, Venetic, and Sabellic is not uncontroversial, though it is supported by substantial phonological and morphological evidence: the merger of **b^h*- and **d^h*- as **f-*,¹¹ the gerundive in **-nd-*, the ipf. subj. **-sē-*, the ipf. **-βā-* (the more probative morphological features are unattested in the fragmentary Venetic corpus). Proto-Italic was recognized as a node from the start of the serious scientific investigation of the Indo-European languages. But some scholars beginning with Walde (1917), Muller (1926), and Devoto (1929) have challenged this assumption and argued instead that Italic and Sabellic are two separate branches that have undergone a secondary process of convergence.¹²

And indeed, there is no doubt that much convergence has happened between Latin and Sabellic. For example, the change of intervocalic **-z-* to *-r-*, called rhotacism, affects both Latin and Umbrian but not Oscan and can be shown to have happened long after the initial separation of both languages. In Latin the change happened sometime in the fourth century, and the Umbrian change may have happened around the same time. Initial *dī-* is eventually simplified in Latin, Umbrian, and Oscan to *i-* (except in Bantia), but again these changes happened within the historical record for Oscan and Latin at least. Both Latin and Sabellic show deletion of the final primary marker **-i* in the 1sg., 2sg., 3sg.,

⁹ On a kylix from Aidone. See Lejeune 1990.

¹⁰ For an ingenious attempt to make sense of this text, see Martzloff 2011.

¹¹ But if, as I argued in Weiss 2018a (following Walde 1906), **d^hrag^h* ‘drag’ > **drag^h* > **trag^h* > *trah-ere* by Limited Latin Grassmann’s Law and the change of **d^h-* to *tr-*, and if Limited Grassmann’s Law is only Latin, then there would be evidence that reflexes of **b^h*- and **d^h*- did not fall together in initial position in Italic so that the *f-* from both **b^h* and **d^h*- would have to be a diffused trait. The problem is that we can’t determine what the date of Limited Grassmann’s Law is, due to the lack of Sabellic evidence. It could be ordered very early in Proto-Italic before the devoicing of voiced aspirates or even theoretically in Proto-Italo-Celtic times.

¹² A thorough survey of the arguments up until 1950 is provided by Diver’s unpublished dissertation of 1953. Further important works are Jones 1951, Safarewicz 1963, Beeler 1966, and Campanile 1968.

and 3pl. (Meiser 1998: 74 after Rix 1996: 158). But the survival of an unapocoped final *-i* in *tremonti* in the *Carmen Saliare* makes it unlikely that this apocope dates to Proto-Italic times. The *Carmen Saliare* is old, but not that old, and the text has a specifically Latin form of the acc. pronoun *tet* and so could not be “Proto-Italic”. Instead, the apocope must be a diffused change.¹³

But while it is easy to show some degree of phonological convergence and, of course, lexical interchange and syntactic influence within the historical period, I know of no case of a Sabellic morpheme being adopted into Latin or vice versa. We have no *v*-perfects in Sabellic, no *-tt*-perfects in Latin, no Latin infinitives in *-om*, no Latin athematic nom.pl. in *-s*, and so on. This difference between phonological, lexical, and syntactic permeability vs. morphological impermeability is not surprising: morphology is known to be more resistant to borrowing, but the absence of morphological borrowing within the attested timeframe, in a period when the Sabellic and Latin languages were in intense contact, should strengthen our confidence in the value of shared morphology for establishing the Proto-Italic subgroup.

There are a number of shared phonological developments unique to the Italic languages that cannot be shown to be the result of convergence and thus are good candidates for defining innovations of Proto-Italic. The difficult issue is deciding whether they are non-trivial. First on this list is the development of the PIE voiced aspirates. In initial position PIE **d^h* and **b^h* developed to *f* and **ġ^h/g^h* to *h* in Latin, Sabellic, and Venetic: **b^huh_x-* > Lat. *fu-ī* ‘I was’, Osc. *fu-st*, Umb. **fu-st** ‘will be’; Transponat **d^hh₁k-* ‘make’ > Lat. *faciō*, Osc. **fakiīad**, Umb. **fačia** subj.3sg., Ven. **vhagsto** pret.3sg.; **ġ^horto-* ‘enclosure’ > Lat. *hortus* ‘garden’, Osc. **hūrz**; Ven. **hosti-** < **g^hosti-* ‘guest’. In medial position, on the other hand, the voiced aspirates became voiced fricatives, and the labial and dental fricatives were not merged:¹⁴ Lat. *nebula* ‘cloud’ < **neβVla* < **neb^hVleh₂*, cf. Ved. *nābhas-* ‘cloud’; *aedēs* ‘temple’ < **h₁aǵd^h-* ‘burn’, cf. Gr. *αἰθόμενος* ‘burning’; Lat. *Samnium* Osc. **safinim** by anaptyxis from **saf-nim* and Gr. *Σάωνιον* point to a Sabellic **saβniō-*;¹⁵ Ven. **loderobos** ‘children’ dat.pl. ← **h₁leud^herob^hos*. The velar fricative went on to become *h* everywhere except Faliscan where it hardened to *g*: **meġ^hei* ‘me’ dat.sg. > Lat. *mihī*, Umb. *mehe*, **leg^heti* > Fal. **lecet** /leget/ ‘lies’, cf. Goth. *ligan*.¹⁶ The combination of devoicing, merger of **d^h* and **b^h* in *f* in initial position and voicing in medial position (whether this voicing directly continues the voicing of the voiced

¹³ For more on diffused changes, see Weiss 2020: 496–7. For what it is worth, Sicel appears to have an unapocoped *esti*.

¹⁴ Though eventually they did merge in Sabellic and Faliscan.

¹⁵ Gr. *Σάωνιον*, which must have been borrowed from a Sabellic form before anaptyxis, shows that the Greeks heard the fricative spelled by the Oscans with **f** as a voiced sound.

¹⁶ The reflex of **g^h* was still an obstruent in Umbrian at the time of medial syllable syncope because when this sound comes together with *t* by syncope it develops to *j* in the same way as *k*, e.g. **ueġ^hetōd* > **ueġetōd* > **arveitu** ‘bring’ parallel to **fakitōd* > **feitu**.

aspirates or is a revoicing) is a set of developments that is found in no other Indo-European branch.¹⁷

A set of sound changes that are certainly shared by Sabellic and Latin is the absorption of a short vowel after yod in a medial syllable. This sound change led to the creation of the 3rd *io*-type in Latin and its analog in Sabellic, e.g. **kap̥iesi* ‘you (sg.) take’ > **kapisi* > *capis*. Cf. Osc. *factud* (Lu 1.9) fut.ipv.sg. ‘make’ < **fak̥itōd* < **fak̥ietōd*. After a base of more than one mora, there was an epenthesis of *i* before the yod prior to absorption of the *e*: **sent-iesi* > **sent-i̯e-si* ‘you (sg.) feel’ > **sent-īsi* > *sent-īs*. Cf. Umb. **amparitu** fut.ipv.sg. ‘raise’ < **am-par-i̯e-tōd*.¹⁸ The sound changes that produced this system appear to be quite early since they predate the resolution of syllabic sonorant consonants (see Fortson 2018), but, on the other hand, these sound changes appear to be distinct from similar changes in Germanic and Celtic.¹⁹

Another interesting phonological development is the outcome of **mmV*. This sequence first arose by the loss of a laryngeal or by Lindeman’s Law. It is also found in the ordinal and superlative suffix **-mmo-*, which is of uncertain analysis. In Latin and Venetic the supporting vowel is *o*, e.g. **g̥h̥mmo* > Lat. *homō* ‘man’, cf. Goth. *guma*, **dek̥mmos* > Ven. *dekomei* loc.sg., cf. Celtib. *tekametam* ‘tenth’. In Sabellic the outcome appears to have originally been *u*. The best evidence for *u* is Osc. **últiumam** ‘last’, Palaeo-Umbrian **setums** (a personal name, lit. ‘seventh’) < **sept̥mmos*, the Pre-Samnite superlative *φολαισμος* ‘best’ nom.pl.²⁰ It’s not certain what the Proto-Italic state was. We can certainly exclude **um* since that would not be lowered to *om* in Latin, cf. *tumor* ‘swollen condition’, *gumia* ‘glutton’. It’s conceivable that **om* would have been raised to *um* in a medial syllable in Sabellic, but there is no independent evidence for such a change. Rather than miss the generalization that the Italic languages uniquely have a rounded vowel as the reflex of

¹⁷ But note that if Sicel is Italic, and if the evidence is correctly interpreted as showing that Sicel had voiced reflexes in initial position, this isogloss would have to be interpreted differently. Either the initial PIE voiced aspirates first became voiced fricatives that were retained or became stops in Sicel and were devoiced in the rest of Italic, or the initial voiced aspirates became voiceless fricatives which were then voiced in Sicel in all positions (cf. the famous southern British English from which the standard dialect borrowed *vat* and *vixen*). Whichever interpretation is correct, the Italic developments would still be unique.

¹⁸ The sound changes that produced the 3rd *io* ~ 4th conjugation contrast are attested outside of verbal morphology, e.g. **d̥i̯eui̯io-* ‘heavenly’ > **d̥i̯ui̯io-* (Osc. **di̯viai̯**), so it is uneconomical to set up athematic *i*-inflection for present stems which are functionally identical to the **je-/i̯o-* presents of other branches.

¹⁹ It is attractive to derive the endings of Old Irish *i*-stem verbs (absolute *gaibid*, conjunct *gaib* ‘takes’) from an immediate preform **-iti* < **-i̯eti*. But there are Gaulish forms that appear to show the unreduced sequence *-ie-* (*bissiet* ‘will be’), and the Sieversish distribution seen in Sabellic does not hold in Old Irish, e.g. *bruinnid*, *bruinn* ‘flows forth’ with an S 2 inflection after a heavy base.

²⁰ The Oscan form **humuns** ‘men’ and Umb. *homonus* dat.pl. are ambiguous. **humuns** is written in an alphabet that does not distinguish *u* and *o* and Umbrian lowers *u* to *o* before *m*.

prevocalic **m̥* which contrasts with its development in preconsonantal position, it may be preferable to reconstruct **om* with a close mid central unrounded vowel that merged with either *o* (in Latin and Venetic) or *u* in Sabellic.²¹ There are a number of other phonological features that could be mentioned, but they are all problematic in one way or another.²² The phonological innovations are admittedly not many, but they are indicative of a subgroup.

It is the shared morphological innovations, which, in my opinion and the opinion of most experts, make the existence of a Proto-Italic unavoidable. The Italic languages share a new verbal adjective, the gerundive, with the suffix *-ndo-* in Latin and **-mmo-* in Sabellic (Osc. **úpsannam** ‘to be constructed’, Umb. *ocrer pihaner* ‘to purify the city’). The origin of this form and the synchronically related gerund, not attested in Sabellic, are much debated. The original function of the form seems to have been quite similar to a middle participle as we can see in synchronically isolated cases such as Lat. *secundus* ‘following’ ~ *sequor* ‘I follow’, *oriundus* ‘arising’ ~ *orior* ‘I rise’, but the semantic development to a verbal adjective of necessity is found in both Sabellic and Latin. Whatever its origin, the gerundive has no analogs outside of Latin and Sabellic.

The imperfect subjunctive in **-sē-*, e.g. Osc. **fusíd** = Lat. ‘foret’, Lat. *es-sē-s* ‘be’ ipf.subj.2sg., is another morpheme of disputed origin.²³ It does not have any *comparanda* outside of Italic.²⁴ The category is not attested in Umbrian or Venetic. But beyond the existence of an identical morpheme for this category, it is worth noting that a subjunctive system with present, imperfect, perfect, and presumably pluperfect is a uniquely Italic way of organizing the verb.

Another Italic-only verbal exponent is the imperfect indicative morpheme **-bā-* (Lat. *dūcēbās* ‘you were leading’, Osc. **fufans** ‘they were’, Vest. *profafa-* i.e. = Lat. *probābā-* ‘was approving’).²⁵ It is generally agreed that **-bā-* is

²¹ Alternatively, one could suppose that pre-vocalic syllabic **m̥* was preserved in Proto-Italic and then developed in slightly different ways in the daughter languages. In this way one loses the generalization that all Italic languages show a rounded vowel in this environment, but it is not too shocking that a prop vowel should develop to a rounded vowel before a labial. A distinct syllabic **m̥* arose in the 1sg. of the verb ‘to be’: **esmi* became **esm̥* by the loss of the final primary marker **-i*, and this developed to *esom* in VOL (Garigliano ESOM), but there is quite a bit of variation here. Latin itself attests *sim*, said to be Augustus’ favored form, and Sabellic has the same two variants **som* (Osc. **súm**) and **sim* (Pre-Samn. and SPic. **sim**). There are also Sabellic forms in **esum** (TE 4 SPic., Ps 4, 5 **esum**) and **sum** (Ps 13), but these are in alphabets that don’t distinguish *u* and *o*.

²² See Weiss 2020: 496–8. On the development of syllabic liquids in Italic, see Zair 2017. Thurneysen-Havet’s Law (in the formulation of Vine 2006) and the development of **CRHC* to *CaraC* are thought to be conditioned by the PIE accent and would be early, potentially of Proto-Italic date. I don’t believe that there are any secure examples of Thurneysen-Havet’s Law in Sabellic, but there is one good instance of **CRHC* to *CaraC* (Umb. *parfa* (type of bird) < **parasā* < **p̥f̥h̥seh₂*, see Höfler 2017). This rule has a close parallel in Greek, however, and it is thus conceivable, though unlikely, that the Latin and Sabellic developments were independent.

²³ Cf. Jasanoff 1991, Meiser 1993, Rasmussen 1996, Christol 2005 for some recent attempts.

²⁴ Despite Campanile’s attempt (1968: 59) to connect it with the Brittonic subjunctive.

²⁵ On the last, see Dupraz 2010: 321.

a form of the root **b^huh_x-* ‘be’ combined with the morpheme **ā < *eh₂* also seen appended to the root **h₁es-* in the unique imperfect stem seen in 2sg. *erā-s* ‘you were’.²⁶ The combination of this extended root shape with a nominal form of the verb, probably originally an instrumental, is only found in Italic. Again, the corresponding forms, if any, in Umbrian and Venetic are unknown.

Another shared Italic innovation is the replacement of the Proto-Indo-European 2nd plural middle ending **-d^hue* with a form containing **-m*: Lat. *-minī*, Sabellic **-mX* inferable from fut.ipv.mid. ending Umb. *-mu* and Osc. *-mur* < Proto-Sabellic **-mōr*.

Ideally, it would be preferable to derive these forms from the same proto-form. If, as most scholars believe, the Latin mid.2pl. ending continues the nom.pl. of a middle participle (< PIE **-mh₁no-*), the Proto-Italic form, if there was one, would have been **-manōs* with the inherited thematic nom.pl. ending. In Latin the analogy that produced the mid.ipv.3sg. was act.ipv.2pl. *-te* : fut.ipv.act.3sg. *-tōd* : mid.ipv.2pl. **-manoj* : fut.ipv.mid.3sg. **-manōd* > *-minō*. That is, the acquirer got the idea that the fut.mid.ipv. was formed by hacking off the final syllable of the act.ipv.2pl. and substituting *-ōd*. But this is not the only way an acquirer could have conceived of the relation. Alternatively the “rule” could have been “remove all material but the initial consonant and substitute with *-ōd*,” i.e. *-te* : *-t-ōd* :: **-manōs* : **-m-ōd*. This path seems the only way to unify the Italic forms and retain the observation that both Sabellic and Latin have replaced the inherited mid.2pl. ending with a form beginning with **-m*.²⁷

There are a few other features shared between Latin and Sabellic, such as the use of the interrogative-indefinite stem as a relative pronoun (Umb. *po-i* = Lat. *quī*). But this is a common development and occurred independently in Hittite, Thessalian, and elsewhere. Another oft-cited commonality is the creation of a distinct ablative singular form for all declensions,²⁸ e.g. Osc. *toutad* ‘community’ from an *ā*-stem. However, Celtiberian also created distinct ablatives for other stem types, e.g. *ā*-stem *arekorataz* (the name of a town attested on a coin, i.e. ‘from the town of A.’).²⁹ Thus this innovation could have happened

²⁶ There are other views, however, e.g. Willi 2016.

²⁷ The alternatives are worse: (1) Latin and Sabellic both replaced the mid.2pl. with *m*-initial forms, which are unrelated. (2) While Latin had a reflex of **-mh₁no-*, Sabellic had **-mo-* like East Baltic and Slavic, but there are to my knowledge no isoglosses connecting Sabellic and Balto-Slavic, and Balto-Slavic **-mo-* may in any case come from **-mh₁no-*. (3) **-mh₁no-* gave **-mo-* in Sabellic, but this is excluded by the many cases of survival of *-mn-*, both primary and by syncope.

²⁸ In the proto-language only *o*-stems made distinct abl.sg. forms. In all other stem types, the abl.sg. and the gen.sg. were identical. Another shared innovation of Sabellic and Latin is the extension of the PIE thematic instrumental to the dative-ablative (VOL *-ois*, Osc. *-ūis*), but Venetic has retained the more archaic *-obos* in *loderobos* ‘children’ dat.pl.

²⁹ See Villar 1995 and Beltrán & Jordán 2019: 251–3. Young Avestan, independently, did the same thing, e.g. *zaoθraīat* ‘libation’ from an *ā*-stem, etc.

(1) independently in Latin, Sabellic, and Celtiberian, (2) in Proto-Italic and Celtiberian, or (3) in Proto-Italo-Celtic.³⁰

The realm of derivational morphology, which is typically underexploited in discussions of subgrouping, also displays a number of striking shared Italic innovations. For example, the suffixes **-āsijō-* (Umb. *farariur* ‘pertaining to grain’ nom.pl.m. = Lat. *farrārius*), and **-āli-* ~ dissimilated to *-āri-* after a base containing an *l* (Umb. *sorsale* ‘of pig’ *staflare* ‘of the stall’ ~ Lat. *mortālis* ‘mortal’, *militāris* ‘military’) are exclusively Italic.³¹ Both Latin and Sabellic have specialized the conglomerate **-kelo-* to form diminutives to nonthematic bases (Osc. *zicolom* Umb. *tiċel* ‘day, date’ ~ Lat *diēcula*). Both Latin and Sabellic have a predominantly deverbal adjectival formant **-d^hli-* (Lat. *amābilis* ‘lovable’, Umb. **purtiŋele** ‘to be offered’).

The shared lexical material of Italic is extensive. Safarewicz (1963) estimated that, with obvious loanwords excluded, 49 percent of the Oscan vocabulary known to him had exact matches in Latin.³² Of course, a true doubter of Italic unity could claim that this high percentage results from borrowing. But there are several items where semantic divergences make recent borrowing unlikely. For example, Lat. *aut* means ‘or’ but in Osc. *avt* means ‘but’. Lat. *enim* means ‘then’, but Osc. **ínim** means ‘and’. We may also note some interesting specializations of meaning and/or form not found outside of Italic: Latin, Sabellic, and Venetic all have the stem **diē-* generalized from the Stang’s Law outcome of **diēum* as the word for ‘day’: Lat. *diēs*, Ven. loc. *diei*, Osc. *zicolom*, Umb. **tiċel** < **diēkelos*. Though the accusative form was obviously inherited (Ved. *dyām*, Phryg. *Tiav*, Gr. *Zḗv(a)*), it is only in the Italic languages that a new paradigm has been formed specifically in this meaning. Only in Latin and Sabellic does **h₂eh₁seh₂* mean ‘altar’ (Lat. *āra*, Umb. **asa** abl.sg., Osc. **aasai** loc.sg.). The Hittite cognate *haššāš* means ‘hearth’ and Ved. *āsa-* m. means ‘ashes’. Latin, Sabellic, and Venetic have created a neo-root **d^heh₁k-* ~ *d^hh₁k-* from a *k*-extended stem originally at home in the active singular of the aorist (Lat. *facere* ‘to make’, Umb. *faċia*, Osc. *fakiiad* pres. subj.3sg., Ven. **vhagsto** pret.3sg., vs. Gr. *ἔθηκε*, Phryg. *addaket* aor.act.3sg. vs. *ἔθετο* aor.mid.3sg.).³³

³⁰ Closely related to the ablative phenomenon is the extension of originally instrumental adverbs in **-eh₁* > **-ē* by *-d-*: OLat. *FACILUMED* ‘very easily’, Osc. *amprufid* ‘improperly’.

³¹ One might add **-id^ho-* if we were sure that the Sabellic place name *Callifae* were to be equated with Lat. *callidus* ‘experienced’ < **‘hardened’* or Lat. *calidus* ‘warm’. But I am skeptical of this because the name is attested only once at Livy 8.25 beside the much better attested *Allifae* and *Allifae* is known to have had a long *i* (*Allifāna* Hor. *S.* 2.8.39, Ital. *Alife*).

³² See also the listing and discussion in Fortson 2017: 843–5.

³³ There are also many items that are attested exclusively in Latin and Sabellic. To mention just two: **kubā-* ‘lie’ (Lat. *cubat*, SPic. **qupāt**) largely replacing **leg^h-* except in Fal. *lecet*, SPic. **veiiat**, and Lat. *lectum* ‘bed’, **fameliĵā* ‘household’ (Lat. *familia*, Umb. **fameŋia**) derived from **famelō-* ‘slave’ (Lat. *famulus*).

All in all, the phonological, morphological, and lexical innovations shared between Latin, Sabellic, and Venetic (when available) are too numerous and integrated to be the result of secondary approximation alone. At the same time, there are quite appreciable differences between the Italic languages. Given the fragmentary state of the Italic languages other than Latin, it is hard to know exactly how different Latin and Sabellic were in, for example, the second century BCE. In my opinion they were synchronically much less closely related or mutually intelligible than the old Germanic languages but more closely related than Old Irish and Middle Welsh or Lithuanian and OCS. From this we infer that there must have been quite a long period of divergence before the forms of Italic began to converge again in the historic period. Whether this inference can be made to correlate with any plausible archaeological or genetic scenario is an open question.

Two final points on the question of Proto-Italic: What would prove that the Sabellic languages and Latin do not belong to the same subgroup? Well, imagine, as a thought experiment, that a scholar claimed that Sabellic and Germanic formed a subgroup within Indo-European. This would be refuted to most people's satisfaction by pointing out that Sabellic and Germanic share no innovations that (1) are not shared with other groups as well and (2) precede the earliest separate innovations in these two groups. If our imaginary proponent of Sabello-Germanic retorted that Proto-Sabello-Germanic should be reconstructed at a stage that could account for both sets of developments, we would respond that such a reconstructed state would be virtually identical to Nuclear Proto-Indo-European and therefore any Nuclear Indo-European language would be derivable from it.

If we return now to reality, can the proponents of the independence of Sabellic and Latin point to any innovations in either language group that make it impossible to derive the other branch from any other common ancestor than the proto-language? There are a few cases where the differing outcomes of the two groups lead to the reconstruction of the PIE state of affairs (e.g. syllabic nasals, labiovelars), but these must be weighed against the instances where this is not the case. Finally, if, on theoretical grounds, we believe that only binary branching is possible,³⁴ denying an Italic subgrouping immediately raises the question of what Sabellic or Latin should be grouped with instead. And when we put the question in those terms, it becomes clear that there is no other branch that could be more closely grouped with either Latin or Sabellic. And thus one would be forced to the position that, despite the evident shared innovations, Latin or Sabellic is more closely related to some language of which there is no trace.

³⁴ See Hale 2007: 238–9 on this point.

8.3 The Internal Structure of Italic

There are actually fairly few innovations on the Latin side that can be shown to encompass all the Latin dialects and to have taken place before the onset of the historical record. In many cases the fragmentary dialects don't preserve the trait in question. For example, the productive *v*-perfect formant is attested in Roman Latin (earliest epigraphical example *PROBAVET* 'approved' from the Egadi rostra dated before 241 BCE; see [Prag 2014](#)) and in Praenestine (*CAILAVIT* 'chiseled') but not in Faliscan. This is presumably an accidental gap. The scant corpus of Faliscan does not preserve any alternative morpheme in the same functional slot. In some cases what would appear to be a defining innovation of Proto-Latin can be seen not to have affected all the dialects or to have been a later diffused change. Roman Latin, for example, has changed medial **β* and **ḍ* to stops *b* and *d*, and **γ* to *h*, whereas Faliscan has *f* for the reflex of medial **β* and **ḍ* and apparently a stop *g* as the reflex of **γ* (Fal. *lecet* 'lies' < **leg^heti*). Thus Proto-Latin must have had fricatives **β*, **ḍ*, and **γ*, and not **b*, **d*, and **h* as Roman Latin. With these cautions in mind, we may point to the following Latin innovations, which are assumed to be Proto-Latin in the absence of evidence to the contrary.

As far as phonology is concerned, few secure innovations define the Latin node; most prehistoric phonological innovations are on the Sabellic side. One Latin innovative feature is the shortening of long vowels before final *-m*, which did not happen in Proto-Italic since South Picene and Oscan preserve distinct reflexes of a long vowel in the genitive plural in **-ōm* and Oscan retained a long vowel in this environment at least in monosyllables ([Weiss 1998](#); [Zair 2016](#): 82). There was eventually such a shortening in polysyllables in the Sabellic languages, but this is an independent change. Another phonological innovation on the Latin side is the contraction of some unlike vowel sequences after the loss of intervocalic yod. Whereas like vowels contracted already in Proto-Italic (e.g. the iterative-causative suffix **-eje-* > **-ee-* > **-ē-* Lat. *ē*, PSab. **ē*), the sequence **āiō* contracts in Latin to *-ō* but remains uncontracted in Sabellic (Lat. *vocō* 'I call' vs. Umb. *subocau* 'I invoke' < **subuokāiō*). Likewise, the sequence **-āiē-* contracts to *ē* in Latin but remains uncontracted or perhaps diphthongizes in Sabellic (*amēs* 'love' subj.2sg. < **amāiēs*, vs. Osc. *deivaid* 'may he swear' < **deiuāiēd*).³⁵

There are other cases where both Latin and Sabellic have innovated in different ways from the Proto-Italic situation. For example, Latin has eliminated the nasal in the sequence **-Vns#*, e.g. in the acc.pl. (VOL *DEIVOS* 'gods' acc.pl.). The same treatment is found in Venetic (Ven. *deivos* 'gods' acc.pl.). In Sabellic, on the other hand, the development of final **-ns* and **-nts* is to *-f*

³⁵ Cf. also the uncontracted form Umb. *ahesnis* 'brazen' abl.pl. < **ajesno-*, although in this case the Latin cognate *aēnus* also remains – surprisingly – uncontracted.

(Umb. **vitluf** ‘calves’ acc.pl., *traf* ‘across’ ~ Lat. *trāns* < **trānts*). The sequence that gave *-nd-* in the Latin gerund(ive) gave *-nn-* in the corresponding Sabellic morpheme (Osc. **úpsannam** ‘to be constructed’; Umb. *pihaner* /*pianner*/ ‘to be purified’ = Lat. *piandī*).

Sabellic, on the other hand, has a number of distinctive phonological innovations. The most salient is the change of the voiced and voiceless labiovelars to labial stops, e.g. **k^wis* ‘who’ > Osc. *pis*, Umb. *pis-i* = Lat. *quis*; **g^wih₃uo-* ‘alive’ > Osc. *bivus* ~ Lat. *vīvī*, **g^wem-* ‘come’ > Umb. *benust* fut.perf.3sg. ~ Lat. *vēnerit*.³⁶ Venetic agrees with Latin in preserving the voiceless labiovelar and turning **g^w* into *w* (**kve** ‘and’, **vivoi** ‘alive’ dat.sg.). Another distinctive feature of Sabellic is across-the-board syncope of a short vowel before a final *-s* (Osc. **húrz** ‘garden’ = Lat. *hortus*, Osc. *bantins* ‘from Bantia’ < **bantīnos*). In Latin and Venetic this type of syncope is more limited and occurs chiefly after *r* (Lat. *sacer* ‘sacred’ < **sakros*,³⁷ Ven. **teuters** ‘public’ < **teuteros*), but not elsewhere. In Sabellic, stops were lenited to fricatives before a dental stop, so **pt* > *ft*, **kt* > *ht* (Osc. *scriftas* ‘written’ nom.pl.f. ~ Lat. *scriptae*, **úhtavis** ‘Octavius’). In Umbrian **ft* > *ht* (*screihtor* ‘written’ nom.pl.n.). The voiced labial and dental fricatives *-β-* and *-ð-*, which occurred in medial position as the reflexes of the PIE voiced aspirates, merged as *β* ⟨*f*⟩ (Osc. **mefiai** ‘in media’, Umb. **rufu** ‘rubrōs’). In Latin and Venetic these are kept separate, ultimately becoming stops in Latin and probably eventually in Venetic too (**mediai** ‘middle’ loc.sg. from **med^hio/ā-*, **louderobos** ‘children’ dat.pl. from **h₁leud^heros*, cf. Gr. *ἐλεύθερος* ‘free’). In initial syllables, syllabic nasals developed to **aM* in Sabellic but to **eM* in Latin (Osc. **fangvam** ‘tongue’ acc.sg. < **d^hŋg^huā*, an-neg. < **ŋ-* vs. OLat. *dingua* < **d^hŋg^huā*, in- < *en-*). But elsewhere the development is to *en* as in Latin (Umb. *desen-duf* ‘twelve’, cf. Lat. *decem* ‘ten’). In Venetic the outcome is *-an* at least in final syllables (**donasan** ‘they gave’ < **-snt*).

Turning to morphology, the innovations are more evenly distributed between Latin and Sabellic.³⁸ Oscan and very probably Umbrian have remade the inherited nom.sg. of *n*-stem nouns with *-ō* as the final-syllable vowel by introducing the *n* from the oblique stem and recharacterizing the nominative with *-s*. The resulting sequence gave *-f*, e.g. Osc. **úittiuf** ‘use’ < **oitiōn-s* vs. Lat. *-iō*, *-iōnis*. In the *-eh₂-* stems, Sabellic retains a contrast between a reflex of **-ā* < **-eh₂* in the nom.sg. (Osc. **víú**, Umb. *Turso* [name of a goddess]) and **-ǎ* in the vocative (Umb. *Tursa*) whereas Latin has a surprising and not satisfactorily explained *-ǎ* in both the nom. and voc.sg. In Sabellic, the proterokinetic

³⁶ The Sabellic treatment of the voiced aspirate labiovelar was to *f* in medial position: **h₁leug^wh₃-* > Umb. **vufu** ‘votive’ ~ Lat. *voveō*. There are no good examples of initial **g^wh₃*, but it would be surprising if it was anything other than *f*.

³⁷ The nom.sg. *SAKROS* is attested in VOL but it is probably an analogical restoration.

³⁸ For a survey of Italic morphology with many references, see Vine 2017.

i-stem gen.sg. ending **-eīs* was generalized to the *o*-stems and consonant stems (**medikeis** ‘meddix’ gen.sg.). In VOL *o*-stems retain *-osīo*, as in VALESIOSIO ‘of Valesios’ (Hom. Gr. *-oio*, Ved. *-asya*, etc.) beside *-ī*, which eventually replaces *-osīo*, and consonant stems retain *-es* and *-os*. But note that in the case of **-eīs*, Sabellic preserves an ending eliminated by Latin. Sabellic also gets rid of the athematic accusative singular ending **-em* < **-m*, replacing it with thematic **-om*. But, on the other hand, Sabellic retains the athematic nom.pl. **-es*, which is syncopated to *-s* (Marruc. *medix* ‘medix’ nom.pl. < **medikes*), whereas Latin has eliminated this ending in favor of *-ēs* < **-ejes* originally from the *-i*-stems. Sabellic also retains the thematic nom.pl. **-ōs* and the *a*-stem nom.pl. **-ās*, which Latin has replaced with pronominal **-oi* > *-ī* and analogical *-ai* > *-ae*, just as in many other Indo-European languages.³⁹ Sabellic has even extended the thematic nom.pl. nominal ending to the nonpersonal pronouns (Osc. **pūs** ‘who’ nom.pl.m. vs. Lat. *quī*). The neut.pl. in Sabellic has generalized the thematic ending *-ā* < **-eh₂* to athematic forms (Umb. **triuper** ‘three times’), but in Latin the generalization has gone the other way with *-a* < **-h₂* in all paradigms.

In pronominal morphology, Latin has extended the accusatives of the singular personal pronouns by *-(V)d* (VOL, Fal., Praen. *med*) whereas Sabellic has used the particle **-om* (OUmb. **mīom**). Sabellic retains the oblique stem formant *-sm-* in the anaphoric and relative-interrogative stem (SPic. **esmín** loc.sg., Umb. **esmik** dat.sg., **pusme** ‘to whom’), which Latin has replaced (*istī eiīei, cui, quō*). Sabellic has an innovative oblique stem of the anaphoric pronoun **eīs-* created by reanalysis of the genitive plural **eīsōm*, e.g. dat.pl. Osc. *eizois*, Umb. **erēr-unt**. Sabellic has a unique proximal deictic stem **eko-/ekso-*. In Oscan these stems are suppletive, with **eko-* forming the nom.-acc. and **ekso-* forming the oblique stem. Umb. has a unitary stem **esso-* < **ekso-*, which may be the older situation. The corresponding Latin proximal deictic is made from a stem **ho-* (Lat. *hic*, Fal. *hec* ‘here’), which may be continued in the Umb. pronominal form *erihont* ‘the same’ nom. sg.m.⁴⁰

In the personal ending of the verb Latin has generalized the thematic 3rd person plural **-ont(i)* to athematic forms (*sont* ‘they are’; exception: opt.3pl. *sient*) whereas Sabellic has extended the range of the ending *-ent* (Osc. **fīe(n)t**), though *-ont* does survive in Pre-Samnite *vuffrōd* ‘they were’. In the primary 2nd plural SPic. has *-tas* (**videtas** ‘you see’), which must be from **-tās* since **-tas* would have syncopated to **-ts*. But Latin has an incompatible *-tis*. Most probably, Proto-Italic had primary 2du. **-tas*, 2pl. **-tes*;⁴¹ secondary 2du. **-tā*, 2pl. **-te*. Latin would

³⁹ Latin does have instances of nom.pl. *-ās*, but these are probably not archaisms. See Weiss 2020: 252.

⁴⁰ On the Sabellic demonstrative system, see Dupraz 2012.

⁴¹ There is little evidence for 2pl. primary **-tes* outside of Italic. The ending **-tes* may itself have been an analogical creation on the model of primary **-me/os*.

have generalized the plural endings while Sabellic generalized the dual endings and leveled the **ā* from the 2du. secondary to the primary dual ending. In 3rd person middle endings, Umbrian appears to have preserved the Plt. situation with primary *-ter* and *-nter* < **-tro* and **-ntro* contrasting with secondary endings in **-tor*, *-ntor* (Umb. **terkantur** ‘let them see’). Oscan has generalized the primary endings, and Latin has generalized the secondary endings. Sabellic also retains *t*-less mid.3sg. forms (Umb. *ier* ‘one goes’, *ferar* ‘one should carry’, Osc. *loufir* ‘or’, lit. ‘(if) it is wished’), which Latin has eliminated without a trace. The mid.2pl. is not attested in Sabellic but is partially inferable on the basis of the deponent future ipv. ending **-mō* (Umb. **persnimu** ‘pray’, Osc. *censamur*) which must have been created like the corresponding Lat. *-minō* on the basis of the 2pl. middle ending. This form therefore began with **m-*. In the endings of the perfect system, Latin has for the most part preferred the endings originating in the PIE perfect (1sg. *-ai*, 2sg. **-istai*, 3sg. *-eit*, 3pl. *-ēre*, all of perfect origin), but has also incorporated some originally aoristic endings (3sg. *-ed*, 3pl. *-ēr-ont*, and *-(er)ont* ← **-ond*, cf. Fal. **fifqo(n)d**). Sabellic has only aoristic endings in the forms we know: 1sg. *-om*, 3pl. *-ens*, Pre-Samn. *-o(v)δ*. The ending *-e* of the perf.2pl. form Pael. *lexe* ‘you (pl.) have read’ is sometimes compared to Ved. perf.2pl. *-a* but, given the overall aoristic provenance of the perfect endings of Sabellic, this is improbable. The ending may ultimately be from **(-s)te*.

When we turn to tense, aspect, and mood, Latin has cobbled together a future tense out of (1) an original periphrastic with **b^huh_x-*, which gives the *b/f*-future (Fal. *carefo* ‘I will lack’, Lat. *carēbō* ‘I will lack’), and (2) the PIE subjunctive (athematic *erō* ‘I will be’, thematic *dūcēs* ‘you will lead’). Sabellic has a probable trace of the thematic subjunctive formant *ē* in SPic. **knúskem** ‘know’ 1sg., but it is difficult to say whether this is used in subjunctive or future function. Otherwise, Sabellic has generalized the athematic *s*-future to all stem types. Latin once had comparable forms, but only the PIE subjunctive and optative of this type survive in the mainly OLat. *faxo* (fut.), *faxim* (subj.) type. In addition, the Latin future perfect and perfect subjunctive (*fēcerō*, *fēcerim*) employ this same morpheme after a union vowel /i/ appended to the perfect stem. Sabellic forms the future perfect with the same athematic *s*-morpheme added to a union vowel *-u-*,⁴² but the perfect subjunctive does not use the *s*-morpheme at all, instead adding *-ē-* to the perfect stem (Osc. **trībaraka-tt-ī-ns**). This *ē* is presumably the same as the subjunctive formant *ē* used in the present stem.

The stem formation of the perfect is very divergent between Latin and Sabellic.⁴³ While both Latin and Sabellic continue reduplicated perfects,

⁴² See on this category most recently Zair 2014.

⁴³ This disagreement contrasts strikingly with the general agreement of present stem formation between Latin and Sabellic. The reasons for this contrast are presumably (1) that the historical “perfect” of Latin and Sabellic results from the parallel independent merger of the PIE perfect

simple perfects (perhaps thematic aorists or dereduplicated perfects), and lengthened-grade perfects, Latin has utilized the *s*-aorist formant much more than Sabellic, which does not have a single completely certain example. Venetic, on the other hand, seems to have generalized the *s*-aorist as the productive perfect formant (**donasto** ‘gave’ 3sg., **donasan** 3pl., **vhagsto** ‘made’ 3sg.). Latin has greatly expanded the range of the *v*-perfect, of which Sabellic has no trace. Sabellic has a number of innovative perfect formants mostly of quite unsettled origin. These are the *-tt*-perfect and the supposed *-k*-perfect of Oscan, the *-nk̑i-* perfect of Umbrian, and the *-ō*-perfect of South Picene.⁴⁴

In nonfinite morphology, Latin and Sabellic have made different choices in grammaticalizing different case forms of different stem types as an infinitive. Latin makes use of the locative of an *s*-stem (*dūcere* ‘to lead’ < **deukesi*) whereas Sabellic used the accusative in *-om* (Osc. **trībarakavūm** ‘to build’), which might originate in either a thematic or a root noun (since Sabellic replaces **-em* with *-om*). For the medio-passive, Sabellic retains the form **-fiē* or **-fiē̑* (Umb. *cehefi* ‘to be taken’, Osc. **sakrafir** ‘to be sanctified’), which is an instrumental or dative of the same piece of nominal morphology that gave Ved. *-dhyāi* and Toch.B, Toch.A *-tsi*. Latin has perhaps redone the expected cognate **-diē* as *-rier* to create its passive infinitive (see Fortson 2012; 2013).

Nominal derivational morphology is overall quite similar between Latin and Sabellic. Most suffixes found in one branch are also found in the other in more or less the same function. One notable difference is that Sabellic has no hesitation in adding the suffix *-ȋo-* to a base in *-ȋo-*. This is the origin of the Sabellic (mainly attested in Oscan) gentilics in **-iis** *-ies* (**statiis** < **stātȋȋ-ȋo-s* derived from the praenomen **staatis** < **stātȋȋ-o-s*). Latin has no trace of such iterative derivation and prefers formations like *Lūcilius* and *Mānilius* from *Lūcius* and *Mānius* or *Lātīnus* from *Lātium*. An interesting mismatch between Latin and Sabellic is shown by Lat. *fānum* ‘shrine’ < **fasnom* vs. Osc. **fīfsnu** nom.sg.f., where Latin continues a zero-grade of the root **d^hh₁s-* and Sabellic reflects a full-grade **d^heh₁s-*. Since ablaut in a *-no-* or *-nā-*stem is unlikely, it is possible that the derivational base showed ablaut or that one or the other forms may have been remade on the basis of related elements.

Defining the distinctive lexicon of Latin vs. Sabellic is challenging given the incommensurate sizes and natures of the corpora. For example, even if we combine all Sabellic languages, we still know fewer than eighty of the 200

and aorist and (2) that denominal verbs had not yet acquired a productive perfect formant. For the development of the perfect system of Italic, see in general Meiser 2003.

⁴⁴ On these formants, see for the most recent proposals and a review of earlier scholarship: Willi 2010; Dupraz 2016: 340 (*-nk̑i-*); Willi 2016; Dupraz 2016: 347 (*-tt-*); Dupraz 2018 (*-k-*); Zair 2014 (*-ō-*).

items on the Swadesh list. It is often impossible to determine whether Latin and Sabellic agree on any particular lexical innovation. Nevertheless, something can be said about the distinctive lexical profile of Sabellic (see Buck 1928: 12–17). There are a small number of cases where Sabellic retains a form of a root or particle that has been completely eliminated from Latin.⁴⁵ Sabellic, like Vedic, has reflexes of both **u_{ih}ro-* (Umb. *ueiro* acc.pl.n.) and **h₂nēr* (SPic. *nīir*, Umb. *nerf* acc.pl. etc.) in the meaning ‘man’ while Latin has eliminated the latter.⁴⁶ The Indo-European word for ‘daughter’ is preserved in Osc. *futīr* while Latin has replaced it with the neologism *fīlia*. Likewise, an old word **putlo-* ‘son’ survives in Osc. *puklum* acc.sg., cf. Ved. *putrá-* in contrast to Lat. *fīlius*. Umbrian preserves the *r/n*-stem for ‘water’ (*utur* acc.sg., *une* abl.sg.) while Latin only continues the derivative *unda* ‘wave’⁴⁷ and has replaced the basic word with a West IE neologism **ak^wā* > Lat. *aqua*.⁴⁸ Sabellic has a word for ‘god’ **aišo-* (Osc. *aisús* nom.pl.) that it shares with Venetic (*aisus*), to the exclusion of Latin. Interesting are the divergent prepositional/preverb forms: Osc. *aa-*, Umb. *aha-* ‘to’ (OHG *uo-*) with no analog in Latin; Osc. *eh*, Umb. *ehe* ‘from’ < **eĝ^h*, which, like Lith. *iž*, preserves an *s*-less form of this particle whereas Latin has only *ex* and its further developments; and **dād* ‘from’ (Osc. *dat*, Umb. *da-*), which has no analog outside of Sabellic at all. Some other Indo-European roots and words are preserved uniquely in the Sabellic branch: **ad-* ‘law’ (Umb. *arsmo* ‘rites’, OIr. *ad* ‘law’, *ada* ‘fitting’), Umb. *e-isicurent* ‘seek’ fut.perf.3pl. (Ved. *icchāti*),⁴⁹ Osc. *cadeis* ‘hostility’ (OHG *haz* ‘hate’), Osc. *mais* ‘more’ (Goth. *mais* ‘more’), **nertero-* ‘left’ (Umb. *nertru*, Osc. *nertrak* ~ ON *norþr* ‘north’), Osc. *nessimas* nom.pl.f. ‘nearest’ (OIr. *nessam* ‘nearest’), Umb. *pir* ‘fire’ (Gr. *πῦρ*, but possibly in Lat. *pūrgāre* ‘to purify’), Umb. *terkantur* ‘look’ subj.3pl. (Gr. *δέρκομαι*), Osc. *touta* ‘people’ (OIr. *túath*, Goth. *þiuda*, etc.). There appears to be no particular pattern to these items. Some have matches only in the Northern European IE languages (**ado-*, **toutā*). Most have widespread cognates.

The syntax of Sabellic and Latin are very similar, but this may be partly the result of the generic similarities of epigraphical documents from Central Italy. The use of the locative case in Latin has been greatly curtailed, but the Sabellic

⁴⁵ I am intentionally omitting any proposal of my own.

⁴⁶ The gens Claudia, of legendary Sabine origin, introduced the praenomen and cognomen *Nerō* ‘strong’ into Latin.

⁴⁷ Unless *unda* is deverbative to **u-ne-d-ti* (Ved. *unāti* ‘flows’).

⁴⁸ Oscan *aapam* acc.sg.fem. is usually compared with Ved. *āp-* ‘water’, but this root is **h₂ep-* and would not have a long *ā* in its paradigm. The Oscan word does not mean ‘water’ per se but most likely ‘water works’ vel sim. and can be explained as the substantivization of an inner-Italic *v̥ddhi* formation **āpo-* ‘of water’, but this could just as well be derived from **apā* < **ak^wā* as from **ap-* < **h₂ep-*.

⁴⁹ Unless the root **h₂eīs-* is somehow continued in Lat. *quaerere* ‘to seek’ and *aeruscāre* ‘to beg’.

languages continue to use it freely: Osc. **eisei terei** ‘in this territory’, *comenei* ‘in the assembly’.

8.4 The Relationship of Italic to the Other Branches

Italic, unsurprisingly, shares many features with other European branches of Indo-European. Meillet (1922) famously recognized a “civilization of the Northwest”, which was shared by the branches that eventually became Balto-Slavic, Germanic, Italic, and Celtic.⁵⁰

One occasionally sees the claim that Italic has an especially close relationship with Germanic (most recently Kuz'menko 2011), but there are in my view no innovations in phonology or inflectional morphology shared exclusively by Italic and Germanic.⁵¹ Most of the innovations listed by proponents of Italo-Germanic such as Hirt 1897 or Devoto 1936 are not correct, are attested elsewhere, or are suspect of being parallel developments.⁵²

That said, there are still a number of unique Italo-Germanic agreements in lexicon. Only in Germanic and Italic is the suffix **-no-* added to the multiplicative adverb **duis* ‘twice’ to create **duisno-* ‘double’, a proto-form continued in the Latin distributive numeral *bīnī* ‘two at a time’ and in Germanic **tūiznā-* (ON *tvennr* ‘two-fold’, OE *ge-twinn* ‘twin’, OHG *zwirnōn* ‘to twine’). But the addition of the suffix **-no-* to adverbial forms is well known elsewhere in Indo-European, cf. Ved. *purāṇā-* ‘ancient’ ← *purā* ‘of old’. Lat. *vadum* ~ ON *vað* n. ‘ford’, OHG *wat*, OE *wæd* ‘sea’ is a perfect and exclusive match. Assuming Germanic **ga-* really is cognate with Latin *com-*, the match between Lat. *commūnis* and Goth. *gamains* etc. is striking. An innovative word for ‘year’, Lat. *annus*, Osc. *akenei* loc. sg., and Goth. *apna-*, is added to PIE **h₁ieh₂r*, which survived in both Italic (*hornus* ‘this year’ < **ho-īVrno-*) and Germanic (Goth. *jer*), but this word is probably closely related to the Celtic words for ‘time’ OIr. *am*, Gaul. *amman* (see Stifter 2017: 220–2). The words for ‘barley’, OHG *gersta* and Lat. *hordeum*, reflect two different genitival derivatives of a **g^hr̥sdo-* ‘prickly plant’ (OE *gorst*). Italic and Germanic share not one but two exclusive words for ‘be silent’: (1) Lat. *tacēre*, Umb. *tačez*, Goth. *þahan*, ON

⁵⁰ For more on the Indo-European of the Northwest see Chapter 7.

⁵¹ Polomé 1966 refutes the supposed morphological and semantic isoglosses well, though obviously we might conduct the refutation somewhat differently today.

⁵² The supposed match between the *-ne* of Lat. *superne* and the *-na* of Goth. *utana* ‘from the outside’ is vitiated by the fact that Latin *superne* has a short final *-e* (cf. also *dōnicum* ‘until’ < **dōnVl^kom*), and the suffix *-nV* added to preverbs is found widely in Indo-European, e.g. Hitt. *istarna* ‘among’. The derivation of adverbs from preverbs and pronominal bases with **-t(e)rō* or **-t(e)rōd* as in Lat. *ultrō* ‘willingly’, Osc. *contrud* ‘contrary to’, and Goth. *innap^{ro}* ‘from within’, etc. is not exclusive (cf. Gr. *πρωτέρω* ‘further’ etc.), and the semantic match between Italic and Gothic is not good. While the Gothic forms have ablative meaning, the same is not true of the Italic forms.

pegja, OHG *dagēn* and (2) Lat. *silēre*, Goth. *anasilan* ‘to grow calm’. Both verbs have been compared more widely (LIV² 495 explains *tacēre* ~ *ḡahan* as a semantic specialization of **pteh₂k-* ‘to cower’; others prefer to connect these forms with OIr. *tachtaid*, MW *tagu* ‘strangle’ < ‘silence’, and *silēre* has been compared to OIr. *silim* ‘pour’ and more generally to the root **seh₁-i-* ‘let’), but even if these somewhat difficult comparisons are correct, the close formal and semantic match between Germanic and Italic remains.

A word needs to be said about two matches between Venetic (considered here as an Italic language) and Germanic. The Venetic accusative of *ego* ‘I’ is *meḡo*, and this has been compared to the Germanic accusative reflected by Goth. *mik*. Given the different extensions of the accusative stem of the 1sg. personal pronoun seen in Latin (OLat. *mēd*), Sabellic (OUmb. *miom*), and Venetic, it seems that Proto-Italic must have actually retained an unextended accusative **mē* or **me*.⁵³ This would mean that *meḡo* would have resulted from a secondary influence of Germanic, but given the fact that *meḡo* can be explained as an inner-Venetic conflation of **egō* and **me*, it’s preferable to leave the Germanic and Venetic forms unconnected.⁵⁴ The second Venetic-Germanic isogloss, Venetic *sselboisselboi* ‘for himself’ ~ Goth. *silba* ‘self’, OE *seolf*, OHG *selbselbo*, ON *sjálfr* is unquestionable and not likely to be coincidental. But it should be noted that the Venetic form occurs on one of the latest Venetic inscriptions, written in the Roman alphabet. It’s possible that this is the result of a quite recent and perhaps one-off borrowing from Germanic.

Although there have been the occasional attempts to connect Italic more closely with one or another branch of Indo-European, these are mainly of historical interest. In the early days of Indo-European, some scholars posited a Greco-Italic group, e.g. Georg Curtius (1858: 22).⁵⁵ In both Greek and Italic the voiced aspirates end up as voiceless segments at least in word-initial position, but the behavior of medial voiced aspirates is quite different.⁵⁶ Greek and Italic share an innovative nominal genitive plural for **eh₂-*stems **-āsōm* (Lat. *-ārum*, Osc. *-azum*, Hom. Aeol. Gr. *-ᾱων*), but this is an introduction of the pronominal form (cf. Ved. *tāsām* ‘of these’) into the nominal paradigm, which could have happened independently. An interesting agreement in derivational morphology is the extension of the stem **deḡs(i)-* ‘on the right’ by the oppositional suffix **-teros* (Lat. *dexter*, Umb. *destram*, Gr. *δεξιτερός*) in contrast to **deḡs(i)-uos* (also in Gr. *δεξι(φ)ός*, Gaul. *dexiva*, Goth. *taihswō* ‘the right hand’) or **deḡs(i)-no-* (Ved. *dákṣiṇa-*, YAv. *dašina-*, Lith. *dėšinas*, OCS *desnъ*). The

⁵³ I’ve suggested (Weiss 2018b: 351) that exactly such a form, *me*, is continued in Venetic *me* in the Isola Vicentina inscription.

⁵⁴ Cf. also Hitt. *ammuk*.⁵⁵ The most recent proponent of this view was Thibau (1964).

⁵⁶ The voiced aspirates were also devoiced in Romani.

antonym *σκαί(φ)ός* ~ *scaevus* ‘left’ is also limited to Greek and Latin. An exclusive match of derivation and meaning in a culturally important word is Lat. *liber* ‘free’, pl. *liberī* ‘children’, Ven. *louderobos* ‘children’ dat.pl., Gr. *ἐλεύθερος* ‘free’. The use of the suffix *-ero- suggests that **h₁leud^heros* originally meant ‘(one who is) in the people’ (**h₁leud^ho-* > ON *ljóðr* ‘people’, OE *lēod* ‘man’, OCz. *ľud* ‘people’) in opposition to those outside of the community.

Martynov (1978) has indicated a number of cases where Slavic has two words for the same thing, one of which closely matches Italic. Most of these comparisons don’t hold up – the two best are OCS *mesęcb* ~ OCS *luna* ‘moon’, cf. Lat. *lūna* and OCS *starъ* ~ OCS *matorъ* ‘old’, cf. Lat. *mātūrus* – but these are not significant enough to support any theory of closer connection between Italic and Slavic, let alone Martynov’s view of a prehistoric conquest of pre-Slavs by pre-Italic speakers.

Finally, Melchert (2016), following in the footsteps of Puhvel (1994), has pointed out a few instances where Latin shares some innovative features with Anatolian, e.g. (1) HLuv. REL-*ipa* ‘indeed’ ~ Lat. *quippe* ‘indeed’ < **k^wid-pe*, (2) Lyd. *nāv, nā-m qid* ‘whatever’ ~ Lat. *quidnam* ‘what on earth?’, (3) Hitt. *kappūwe/a-* ‘count’ ~ Lat. *computāre* ‘to reckon’. Whether these agreements are sufficient to support some secondary contact between Proto-Italic and Proto-Anatolian is uncertain.

To sum up: the similarities noted between Italic and other Indo-European branches are predominantly lexical, and, when we compare these similarities to the ones noted between Italic and Celtic, the case for Proto-Italo-Celtic seems all the stronger (Chapter 7).

8.5 The Position of Italic

See Chapter 7.

References

- Adiego, Ignasi-Xavier. 2015. Some remarks on the new Opic (“Pre-Samnite”) inscription of Niumsis Tanunis. *Incontri Linguistici* 38. 125–38.
- Agostiniani, Luciano. 1992. Les parlers indigènes de la Sicile prégréceque. *LALIES. Actes des sessions de linguistique et de littérature* 11. 125–57.
- Ambrosini, Riccardo. 1984. Lingue nelle Sicilia pregreca. In Adriana Quattordio Moreschini (ed.), *Tre millenni di storia linguistica della Sicilia*, 13–36. Pisa: Giardini.
- Bakkum, Gabriël C. L. M. 2009. *The Latin dialect of the Ager Faliscus: 150 years of scholarship. Part 1*. Amsterdam: Amsterdam University Press.
- Beeler, Madison. 1966. The interrelationships within Italic. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 51–8. Berkeley: University of California Press.

- Beltrán, F. & C. Jordán. 2019. Writing and language in Celtiberia. In Alejandro G. Sinner & Javier Velaza (eds.), *Palaeohispanic languages and epigraphies*, 240–303. New York: Oxford University Press.
- Buck, Carl Darling. 1928. *A grammar of Oscan and Umbrian*. 2nd ed. Boston: Ginn.
- Campanile, Enrico. 1968. *Studi sulla posizione dialettale del Latino*. Pisa: Pacini Mariotti.
- Campanile, Enrico. 1969. Note sulle glosse sicule e sui rapporti linguistici fra siculo e latino. In *Studia Classica et Orientalia Antonino Pagliaro oblata*. Vol. 1, 293–322. Rome: Herder.
- Christol, Alain. 2005. Subjonctif latin (-sē) et futur indien (-sya-). *Journal of Latin Linguistics* 9. 25–36.
- Clackson, James. 2015. Subgrouping in the Sabellian branch of Indo-European. *Transactions of the Philological Society* 113. 4–37.
- Curtius, Georg. 1858. *Grundzüge der griechischen Etymologie. Erster Theil*. Leipzig: Teubner.
- Devoto, Giacomo. 1929. Italo-greco e italo-celtico. *Archivio glottologico italiano* 22/23. 200–40.
- Devoto, Giacomo. 1936. Germanisch-Lateinisch und Germanisch-oskisch-umbrisch. In *Germanen und Indogermanen: Festschrift für Herman Hirt*. Vol. 2, 533–47. Heidelberg: Winter.
- Diver, William G. 1953. *The relation of Latin to Oscan-Umbrian*. PhD thesis, Columbia University.
- Dupraz, Emmanuel. 2010. *Les Vestins à l'époque tardo-républicaine: Du nord-osque au latin*. Mont-Saint-Aignan Cedex: Publications des universités de Rouen et du Havre.
- Dupraz, Emmanuel. 2012. *Sabellian demonstratives: Forms and functions*. Leiden: Brill.
- Dupraz, Emmanuel. 2016. Zu einigen Perfektbildungen im Sabellischen. *Indogermanische Forschungen* 121. 333–63.
- Dupraz, Emmanuel. 2018. Les soi-disant parfaits en -k- de l'osque: entre épigraphie, histoire des religions et linguistique comparative. *Les études classiques* 86. 49–67.
- Fortson, Benjamin W., IV. 2012. Latin -rier and its Indo-Iranian congeners. *Indogermanische Forschungen* 117. 75–118.
- Fortson, Benjamin W., IV. 2013. Pre-Italic *-d^hiē̯ (*-dh^heh₁) versus pre-Indo-Iranian *-d^hiōi: Bridging the gap. In Adam I. Cooper, Jeremy Rau & Michael Weiss (eds.), *Multi nominis grammaticus: Studies in Classical and Indo-European linguistics in honor of Alan J. Nussbaum on the occasion of his sixty-fifth birthday*, 50–60. Ann Arbor, MI: Beech Stave.
- Fortson, Benjamin W., IV. 2017. The dialectology of Italic. In Jared S. Klein, Brian D. Joseph, & Matthias Fritz (eds.), *Handbook of historical and comparative Indo-European linguistics*. Vol. 2, 835–57. Berlin: De Gruyter Mouton.
- Fortson, Benjamin W., IV. 2018. A cover-up: Latin amiciō. In Olav Hackstein & Andreas Opfermann (eds.), *Priscis libentius et liberius novis: Indogermanische und sprachwissenschaftliche Studien. Festschrift für Gerhard Meiser zum 65. Geburtstag*, 49–64. Hamburg: Barr.
- Franchi de Bellis, A., 2005. *Iscrizioni prenestine su specchi e ciste*. Alessandria: Edizioni dell'Orso.
- Hale, Mark. 2007. *Historical linguistics: Theory and method*. Malden, MA: Blackwell.

- Hartmann, Markus. 2018. Sicilian. In Jared S. Klein, Brian D. Joseph, & Matthias Fritz (eds.), *Handbook of historical and comparative Indo-European linguistics*. Vol. 3, 1854–7. Berlin: De Gruyter Mouton.
- Hirt, Herman. 1897. Die Stellung des Germanischen im Kreise der verwandten Sprachen. *Zeitschrift für deutsche Philologie* 29. 289–305.
- Höfler, Stefan. 2017. Observations on the *palma* rule. *Pallas: Revue d'études antiques* 103. 15–23.
- Jasanoff, Jay H. 1991. The origin of the Latin imperfect subjunctive. *Historische Sprachforschung* 104. 84–105.
- Kuz'menko, Ju. K. 2011. O čem mogut svidetel'stvovat' germano-italijskie innovacii [What can Germano-Italic innovations tell us?]. *Acta Linguistica Petropolitana* 7. 9–30.
- Lejeune, Michel. 1990. Notes de linguistique italique XL: 'Bois!' disait ce Sicule; 'je boirai' répond ce Falisque. *Revue des études latines* 68. 28–30.
- Machajdík, Barbora. 2018. Dve grécke výpožičky na sikulských nápisoch (*darnakei, poterom*) [Two Greek loanwords in Sicel inscriptions (*darnakei, poterom*)]. *Hortus Graeco-Latinus Cassoviensis* 2. 143–54.
- Martynov, V. V. 1978. *Balto-slavjano-italijskie izoglossy: Leksičeskaja sinonimija* [Balto-Slavo-Italic isoglosses: Lexical synonymy]. Minsk: Nauka i technika.
- Martzloff, Vincent. 2011. Variation linguistique et exégèse paléo-italique: l'idiome sicule de Montagna di Marzo. In Gilles Van Heems (ed.), *La variation linguistique dans les langues de l'Italie préromaine*, 93–129. Lyon: Maison de l'Orient et de la Méditerranée.
- Meillet, Antoine. 1922. *Les dialectes indo-européens*. 2nd ed. Paris: Champion.
- Meiser, Gerhard. 1993. Uritalische Modusyntax: zur Genese des Konjunktiv Imperfekt. In H. Rix (ed.), *Oskisch-Umbrisch: Texte und Grammatik*. 167–95. Wiesbaden: Reichert.
- Meiser, Gerhard. 1998. *Historische Laut- und Formenlehre der lateinischen Sprache*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Meiser, Gerhard. 2003. *Veni vidi vici: Die Vorgeschichte des lateinischen Perfektsystems*. Munich: Beck.
- Melchert, H. Craig. 2016. "Western affinities" of Anatolian. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 297–305. Wiesbaden: Reichert.
- Muller, Frederick. 1926. *Altitalisches Wörterbuch*. Göttingen: Vandenhoeck & Ruprecht.
- Poccetti, Paolo. 2012. Language relations in Sicily. In Olga Tribulato (ed.), *Language and linguistic contact in ancient Sicily*, 49–94. Cambridge: Cambridge University Press.
- Poccetti, Paolo. 2017. The documentation of Italic. In Jared S. Klein, Brian D. Joseph, & Matthias Fritz (eds.), *Handbook of historical and comparative Indo-European linguistics*. Vol. 2, 733–43. Berlin: De Gruyter Mouton.
- Polomé, Edgar. 1966. Germanisch und Italisch im Lichte der deskriptiven Sprachbetrachtung. *Orbis* 15. 190–9.
- Prag, Jonathan R. W. 2014. Bronze rostra from the Egadi Islands off NW Sicily: The Latin inscriptions. *Journal of Roman Archaeology* 27. 33–59.
- Puhvel, Jaan. 1994. West-Indo-European affinities of Anatolian. In George Dunkel et al. (eds.), *Früh-, Mittel-, Spätindogermanisch*, 315–24. Wiesbaden: Reichert.

- Rasmussen, Jens Elmegård. 1996. On the origin of the Latin imperfect subjunctive. *Copenhagen Working Papers in Linguistics* 4. 143–8.
- Rix, Helmut. 1996. Review of Peter Schrijver, *The reflexes of the Proto-Indo-European laryngeals in Latin*, Amsterdam: Rodopi, 1991. *Kratylos* 41. 153–63.
- Safarewicz, Jan. 1963. Two Italic studies. In Jan Safarewicz, *Linguistic studies*, 121–49. The Hague: Mouton. Originally published as “Dwa studia italskie” in *Zeszyty naukowe Uniwersytetu Jagiellońskiego* 60, *Prace językoznawcze* 5, 7–34.
- Stifter, David. 2017. Varia II: The origin of time. *Ériu* 67. 219–26.
- Thibau, R. 1964. *Les rapports entre le latin et le grec: Essai de réhabilitation de l'hypothèse italo-grecque*. Leiden: Brill.
- Villar, F. 1995. *A new interpretation of Celtiberian grammar*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Vine, Brent. 2006. On “Thurneysen-Havet’s Law” in Latin and Italic. *Historische Sprachforschung* 119. 211–49.
- Vine, Brent. 2017. The morphology of Italic. In Jared S. Klein, Brian D. Joseph, & Matthias Fritz (eds.), *Handbook of historical and comparative Indo-European linguistics*. Vol. 2, 751–804. Berlin: De Gruyter Mouton.
- Walde, Alois. 1906. Aspiratendissimilation im Latein. *Indogermanische Forschungen* 19. 98–111.
- Walde, Alois. 1917. *Über älteste sprachliche Beziehung zwischen Kelten und Italikern*. Innsbruck: Wagner.
- Weiss, Michael. 1998. On some problems of final syllables in South Picene. In Jay H. Jasanoff, H. Craig Melchert & Lisi Oliver (eds.), 1998. *Mir curad: Studies in honor of Calvert Watkins*, 703–15. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Weiss, Michael. 2010. *Language and ritual in Sabellic Italy*. Leiden: Brill.
- Weiss, Michael. 2018a. Limited Latin Grassmann’s Law: Do we need it? In Dieter Gunkel et al. (eds.), *Vina diem celebrent: Studies in linguistics and philology in honor of Brent Vine*, 438–47. Ann Arbor, MI: Beech Stave.
- Weiss, Michael. 2018b. *Veneti* or *Venetes*? Observations on a widespread Indo-European tribal name. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 349–58. Ann Arbor, MI: Beech Stave.
- Weiss, Michael. 2020. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor, MI: Beech Stave.
- Willi, Andreas. 2008. *Sikelismos. Sprache, Literatur und Gesellschaft im griechischen Sizilien (8.–5. Jh. v. Chr.)*. Basel: Schwabe.
- Willi, Andreas. 2010. The Umbrian perfect in **-nĕ/-nĕ-**. *Transactions of the Philological Society* 108. 1–14.
- Willi, Andreas. 2016. The Oscan perfect in **-tt-**. *Transactions of the Philological Society* 114. 75–94.
- Zair, Nicholas. 2014. The future perfect in Oscan and Umbrian and the *ō*-perfect in South Picene. *Transactions of the Philological Society* 112. 367–85.
- Zair, Nicholas. 2016. *Oscan in the Greek alphabet*. Cambridge: Cambridge University Press.
- Zair, Nicholas. 2017. The origins of **-urC-** for expected **-orC-** in Latin. *Glotta* 93. 255–89.
- Zamponi, Raoul. 2021. *South Picene*. Abingdon: Routledge.

Anders Richardt Jørgensen

9.1 Introduction

This chapter provides an outline of the defining characteristics of the Celtic proto-language and the internal divisions within Celtic. Only languages which are clearly identifiable as Celtic will be included in this treatment, i.e. Goidelic, Brittonic, Gaulish (including Cisalpine, Transalpine and the onomastic material from Central European and Balkanic Celtic), Celtiberian, Lepontic and Galatian. Pictish, Tartessian and Lusitanian will be excluded, either due to the fragmentary attestation or because it is highly unlikely that the language belongs to the Celtic branch of Indo-European.

9.2 Evidence for the Celtic Branch

When listing the defining innovations of Proto-Celtic, we quickly encounter a problem closely linked to the poor attestation of the Continental Celtic languages: many of the most distinct innovatory features differentiating Celtic from the other Indo-European branches can strictly speaking only be proven to be “Proto-Goidelo-Brittonic”, and it is unclear how close this actually takes us to a Proto-Celtic encompassing both the Insular and the Continental Celtic branches.¹ However, an area where the scant attestation of Continental Celtic nonetheless provides a fair amount of information is historical phonology. Accordingly, Proto-Celtic will mainly be defined by a series of phonological changes differentiating it from Proto-Indo-European and the other Indo-European branches. This does not mean that Proto-Celtic had not innovated in other areas such as morphology and syntax, only that our limited knowledge of Continental Celtic, particularly in the area of verbal morphology, makes it difficult to project innovations such as the *t*-preterite, the *s*-preterite

Work on this chapter was carried out with the support of the Independent Research Fund Denmark for the project *Connecting the dots: Reconfiguring the Indo-European family tree*, and of Riksbankens jubileumsfond for the project *Languages and myths of prehistory: Unravelling the speech and beliefs of the unwritten past (LAMP)*.

¹ Insular and Continental Celtic are used here as geographical terms, without necessarily signifying linguistic subgroups.

and the $\bar{a}$ -preterite back to a specific stage beyond “Insular Celtic” (or Goidelo-Gallo-Brittonic for that matter). For instance, it is not absolutely certain whether the merger of the PIE aorist and the perfect into a new preterite, though completely carried through in Insular Celtic, had necessarily occurred by the Proto-Celtic stage.

In the following, some of the more significant innovations from PIE to Proto-Celtic will be listed in rough chronological order, to the extent that such an order can be established. For more detailed descriptions, see e.g. McCone 1996: 37–104 and Stifter 2017.

9.2.1 *The Centum Merger and “Thorn” Clusters*

The *centum* merger, i.e. the merger of palatal and plain velars, is unconditioned in Celtic and can therefore not be placed with confidence in the relative chronology. Given the equally unconditioned developments in several other Indo-European branches (most notably the neighbouring Germanic and Italic branches), it is likely that this is a very early areal innovation.

Proto-Indo-European sequences of original palatal stop + $*u$ merge with the corresponding labiovelar in Celtic: $*h_1ékwo-$ ‘horse’ > PCelt. $*ek^w o-$ (cf. the Gaul. theonym *Epona*, OIr. *ech* ‘horse’, MW *ebawl* ‘foal’) has the same medial phoneme as PIE $*tek^w-$ ‘runs’ > PCelt. $*tek^w-e/o-$ (OIr. *techid* ‘flees’, MW *tebed* ‘retreat, flight’).

Proto-Indo-European “thorn” clusters, traditionally reconstructed as PIE $*Kp/$ $Gð$ but in fact rather PIE $*TK$, underwent metathesis to $*KT$ in pre-Proto-Celtic, as exemplified by $*h_2r̥tko-$ > $*h_2rkto-$ > $*arxto-$ > PCelt. $*arto-$ (W *arth* ‘bear’, OIr. *art* ‘hero’) and $*d^hǵ^hom-$ ‘earth’ > $*gdom-jo-$ > PCelt. $*gdon-jo-$ ‘earthly; mortal’ (Cisalp. Gaul. *TEUO-XTONI[O]N* ‘of gods and mortals’; simplified to $*don-jo-$ in later Celtic, e.g. OIr. *duine*, MW *dyn*, MBret. *den* ‘man’).

9.2.2 *The Syllabic Liquids*

The Proto-Indo-European syllabic liquids developed a prop vowel whose distribution is mainly conditioned by the following segment. The commonly accepted distribution assumes the outcome $*ri/li$ before stops and $*m$ and the outcome $*ar/al$ elsewhere: PIE $*b^hǵ^h-$ > PCelt. $*brig-$ (Gaul. *-briga*, OIr. *brí*, MW *bre*, MBret. *bre* ‘hill’), $*k^wrimi-$ > PCelt. $*k^wrimi-$ (OIr. *cruim*, MW *pryf*, MBret. *preff* ‘worm’), $*k̥r-n-$ > PCelt. $*karnV-$ (Galat. *κάρον* ‘horn, trumpet’, MW *carn* ‘horn, hoof’, ModBret. *karn* ‘hoof’), $*krso-$ > $*karso-$ > PCelt. $*karro-$ (OIr. *carr*, MW *kar*, MBret. *carr* ‘cart’), $*pr̥h_2-i$ > PCelt. $*(\phi)are$ (Gaul. *are-*, OIr. *air-*, MW *ar-*). This distribution has recently been challenged by Hill (2012), who assumes that $*r/l$ also gave $*ri/li$ before $*n$. This would provide a straightforward explanation for a form

such as OIr. *tlenaid* ‘steals’ < PCelt. **tli-na-* < PIE **tl̥n-ah₂-* from the root **telh₂-* (LIV² 622), which otherwise is difficult to explain satisfactorily. The apparent counterexample PCelt. **karnV-* may be derived from PIE **k̑r̥-snV-* instead.

The differing treatment of PIE **h₂rt̥ko-* ‘bear’ and **h₁r̥ǵʰ-* ‘mount, go up’ in Celtic, PCelt. **arto-* (MW *arth* ‘bear’, OIr. *art* ‘hero’) and **rig-* (OIr. fut. *-rega* ‘will go’, cf. McCone 1996: 62) respectively indicates that, in initial position at least, a preceding **h₂* caused the prop vowel to develop before the syllabic liquid and not after it, as would be otherwise expected. This means that **h₂* still contrasted with **h₁* when the prop vowels developed.

9.2.3 Elimination of the Laryngeals

As is usually the case in non-Anatolian branches, the PIE laryngeals were eliminated as phonemes, but left traces in various ways. Word-initial laryngeals were lost without a trace, whether prevocalic or preconsonantal, while post-vocalic laryngeals in the syllabic coda were lost with compensatory lengthening of the preceding vowel. The latter development took place before the restructuring of the long vowel system outlined below. In a fair number of instances, however, the expected lengthening does not appear, and we are instead left with a short vowel, e.g. PIE **u̯iH-ró-* > PCelt. **u̯iro-* ‘man’ (OIr. *fer*, MW *gwr*), PIE **ǵʰuH-tu-* > PCelt. **gutu-* (OIr. *guth* ‘voice’). This phenomenon, called Dybo’s Shortening (after its first formulation in Dybo 1961), is not restricted to Celtic but is also found in Germanic and Italic, possibly as part of an early areal tendency. The exact conditions leading to this shortening (or lack of lengthening) are not clear, and no consensus has formed as yet. For a recent discussion of the literature on this problem and the proposed solutions, see Zair 2012: 132–50.

Laryngeals between non-syllabic consonants are usually vocalized to **a*, as in e.g. PIE **ph₂ter-* > PCelt. **(ǵ)ater-* (OIr. *athair* ‘father’), PIE **sh₁-tV-* > PCelt. **satV-* (MW *had*, MBret. *hat* ‘seeds’), PIE **p̑l̥h₂-no-* > PCelt. **(ǵ)litano-* (OIr. *lethan*, MW *llydan*, MBret. *ledan* ‘broad, wide’). This appears to be the case irrespective of the position of the syllable in the word, agreeing with Italic but differing from Germanic and Balto-Slavic, where only laryngeals in the first syllable appear to be vocalized to **a*.

Sequences of *CRHC* usually develop into *CRāC* as in Italic, e.g. PIE **p̑l̥h₂-mah₂* > PCelt. **(ǵ)lāmā* (OIr. *lām*, MW *llaw* ‘hand’), PIE **m̑h₂tV-* > PCelt. **mlātV-* (OIr. *mláith* ‘smooth’, MW *blawd* ‘flour’, MBret. *bleut*), but occasionally, a short vowel is encountered instead, e.g. **pȓH-ti-* > PCelt. **(ǵ)rati-* (Gaul. *ratis*, Mlr. *raith*, MBret. *raden* ‘ferns’; cf. Schumacher 2004: 136–7). For recent treatments of the problems relating to the development of laryngeals in Celtic,

cf. McCone 1996: 51–4; Schumacher 2004: 135–8; Zair 2012; Stifter 2017: 1194–6.

9.2.4 *The Syllabic Nasals*

The development of the syllabic nasals is straightforward. As has been demonstrated convincingly by McCone (1992: 21–6; 1996: 70–9; for the traditional view, cf. e.g. de Bernardo Stempel 1987), apparent cases of older **eN* in Irish from PIE **N̥* may effortlessly have passed through the PCelt. stage **aN*, only to have been secondarily raised in the prehistory of Irish. Hence, we may reconstruct PCelt. **aN* as the regular outcome of PIE syllabic nasals in all instances. This is borne out by e.g. Celtib. *argato-* /arga(n)to-/, Gaul. *arganto-*, OIr. *argat*, MW *aryant*, MBret. *archant* ‘silver’ < PCelt. **arganto-* < PIE **h₂(a)rǵ-nt-o-* and Celtib. *tekam-etinas*, Gaul. *dekam-etos* ‘tenth’, OIr. *deich* ‘ten’ (< **deken*) < PCelt. **dekam* < PIE **dekm̥(t)*.

9.2.5 *The Voiced Labiovelar and the Merger of Aspirated and Plain Voiced Stops*

Based on MW *gieu* ‘sinews, tendons’, OCor. *goiu-en* < Brit. plural **gi.ou* (with a secondary *u*-stem plural ending **-ou* < PCelt. nom.pl. **-oues*) < PCelt. **g(i)ṽ-* < PIE **g^w(i)ṽh₂-* (cf. Ved. *jyā-* ‘tendon, string (esp. of a bow)’, Lith. *gijā* ‘thread’, Gr. *βίος* ‘bow’) and Mlr. *nigid* ‘washes’ < PCelt. **nig-i/ṽo-* < **nig^w-ṽe/o-* (Gr. *νίζω*) it appears that PIE **g^w* was delabialized to **g* before a following **i*. For purely structural reasons we would expect PIE **k^w* and **g^{wh}* to be similarly delabialized, but there are no certain instances of this. The delabialization must precede the shift of PIE **g^w* > **b* and consequently the merger of the PIE voiced and voiced aspirate stops (since **g^{wh}* does *not* give PCelt. **b*, but rather PCelt. **g^w*). Therefore, it can safely be ascribed to the pre-Proto-Celtic period, even without any evidence of the sound change from Continental Celtic. In all other instances, when PIE **g^w* was not affected by delabialization, it yielded PCelt. **b* and as such merged with the outcome of PIE **b^h* and the much rarer **b*. This is demonstrated by e.g. Gaul. *-bena*, OIr. *ben*, MCor. *ben-en* ‘woman’ < PCelt. **benā* < PIE **g^wen-h₂* ‘woman’, OIr. *biur*, MW *ber* ‘spear’ < PCelt. **beru-* < **g^weru-* and OIr. *brao*, MW *breuan*, MBret. *brou*, *breau* ‘hand-mill, quern’ < PCelt. **brāuū*, **-on-* < PIE **g^wh₂-u-on-* or **g^wrah₂-u-on-*.

At some point after the development of PIE **g^w* to **b*, the PIE voiced aspirated stops lost their aspiration and merged with the corresponding voiced stops, e.g. PIE **b^hed^h(h₂)-o-* (cf. Lat. *fodiō*, *-ere* ‘to dig’) > PCelt. **bedo-* ‘grave’ (Celtib. *argato-bezom* ‘silver-mine (?)’, MW *bedd*, MBret. *bez* ‘grave’), PIE **seǵ^h-etlo-* (Gr. *ἐχέτηλη* ‘plough-handle’) > PCelt.

**segetlo-* (ModW *haeddel*, MBret. *haezl* ‘plough-handle’, MCorn. *hethlor* ‘ploughman’).

9.2.6 Changes to the Vowel System

The long vowel system was restructured, whereby the PIE long vowel phonemes **ē* and **ō* were eliminated. It is likely that this development had already occurred in the pre-Proto-Celtic period:

- PIE **ō* (including PIE **oH*) was eliminated, giving either PCelt. **ū* (in word-final syllables) or PCelt. **ā* (elsewhere). Accordingly, it merged either with the reflexes of PIE **ū*, **uH* or **ā*, **aH*: Celtib. *n*-stem nom.sg. *-u*, Gaul. *-u*, OIr. *aub* ‘river’ (< **abū* with *u*-infection) < PCelt. **ū* < PIE **ō*, Celtib. *o*-stem Dg. *-ui*, Lep. *-ui*, Gaul. *-ui* < PCelt. **ū* < PIE **ō*, Celtib. sg. *-uz* < PCelt. **ūd* < PIE **ōd*; as opposed to **ā* < **oH* in non-final syllables in e.g. OIr. *dán*, MW *dawn* ‘gift, endowment’ < PCelt. **dānV-* < **doh₃no-* (cf. Ved. *dāná-*, Lat. *dōnum*).
- PIE **ē* (including PIE **eh₁*) was raised to **ī* and merged with the reflex of PIE **ī* and **iH*: Celtib. *ti-*, Gaul. *di-*, MW pref. *di-* < PCelt. **dī* < PIE **deh₁* (Lat. *dē*); the Gaul. onomastic element *-rix* /-rīxs/ ‘king’ in e.g. *Dumnorix*, *Vercingetorix*, OIr. *rí*, gen.sg. *ríg*, MW *rhi* < PCelt. **rīx-s*, gen.sg. **rīg-os* < PIE *(*h₃*)*rēg-*.

This resulted in a triangular long-vowel system, *ī* – *ā* – *ū*. This system was extended with a new *ē* < **e₂* and, somewhat later, a new *ō* < **o₂* during the attested history of the Continental Celtic languages. The Insular Celtic languages may be derived from a long-vowel system with five vowel qualities, *ī* – *ē* (< **e₂*) – *ā* – *ō* (< **o₂*) – *ū*, matching the five short-vowel qualities, *i* – *e* – *a* – *o* – *u*.

Joseph’s Law, formulated by Lionel Joseph (1982; cf. Schrijver 1995: 73–93), states that a pre-PCelt. sequence **eRa* (typically from PIE **eRə*) gives **aRa*. This elegantly explains numerous forms in Goidelic, Brittonic and Gaulish, e.g. **taratro-* ‘drill’ (Ir. *tarathar*, MBret. *tarazr*, Gaul. **taratro-* ⇒ Judeo-Fr. *taredre* /taˈrɛdrə/, OOcc. *taraire*) < **teratro-* < PIE **terh₁-tro-* (cf. Gr. *τέρετρον*) and **talamū* (OIr. *talam* ‘the earth, the world’) < **telamū* < **telh₂-mō*, *-mon-* (Gr. *τελαμών* ‘carrying strap’), **garano-* ‘crane’ (Gaul. *trigaranus* ‘having three cranes’, MW *garan*) < **gerano-* < PIE **gerh₂no-* (Gr. *γέρανος*) which previously had to be reconstructed as **trh₂atro-*, **tlh₂amon-* and **grh₂ano-*. The absence of any traces of Joseph’s Law in e.g. fem. *ā*-stems and weak *ā*-verbs indicates that it was not triggered by a long **ā*. It is also likely that it did not operate when the **a* was word final.

An expanded version of Joseph’s Law has recently been proposed by Eugen Hill (2012). According to Hill, sequences of **eLN*a were also affected by this change. This explains the vocalism of e.g. W *sarnu* as opposed to OIr. *sernaid*

as deriving from a paradigm PCelt. **sternū*, **starnati* (vel sim.), from an older subjunctive **ster-nh₂-e/o-* (cf. Lat. *sternō*).

9.2.7 Cluster Simplification

It is likely that there was a general loss of stops before *-sC-* (Stifter 2017: 1191–2). This would explain instances such as OIr. *tes*, MW *tes* ‘heat’ < PCelt. **testu-* < PIE (Transponat) **tep-s-tu-* and OIr. *lesc*, MW *llesc* ‘weak; lazy’ < PCelt. **lesko-* < PIE (Transponat) **leg^h-sko-* ‘lying down’. In these cases, the root-final stop was not restored, because the association to the underlying root was not sufficiently strong. However, when the association with forms with a preserved root final consonant was sufficiently strong, the consonant was typically restored. The restored stop was subsequently subject to the general neutralization of non-dental stops: before a following **t* or **s* all Proto-Indo-European “non-dental” stops (i.e. **k̂*, **ĝ*, **ĝ^h*, **k*, **g*, **g^h*, **k^w*, **g^w*, **g^{wh}*, **p*, **b*, **b^h*) merge as a velar (or uvular) fricative, usually noted **x*, as in PIE **h₃oktoH* > PCelt. **oxtū* (Gaul. *oxumetos* ‘eighth’, OIr. *ocht*, MW *wyth*, MBret. *eiz* ‘eight’), **septm* > PCelt. **sextam* (Gaul. *sextametos* ‘seventh’, OIr. *secht*, MW *seith*, MBret. *seiz*, PIE (Transponat) **(H)eup-s-elo-* (cf. Gr. *ὕψηλός*) > PCelt. **ouxselo-* (Gaul. *uxello-*, OIr. *úasal*, MW *uchel*, MBret. *uhel* ‘high’). The exact phonemic status of this **x* is not entirely clear; it did not occur in other contexts.

In restored clusters of the structure **xsC*, the **s* was subsequently lost, probably by regular reduction. This paved the way for the *t*-aorist of roots ending in velar stops (as in OIr. pret. *-acht* ‘drove’, MW *aeth*, MBret. *aez* ‘went’ < **ax-t-* < **ax-s-t* from the PCelt. pres. **ag-e/o-*), ultimately deriving from an old 3sg. *s*-aorist. A very similar loss, of both **s* and **x* is observable between liquid and stop. This can be observed in e.g. OIr. *tart* ‘thirst’, MW *tarth* ‘steam, haze’ < PCelt. **tartu-* < **tarstu-* < **trs-tu-* and MW *arth* ‘bear’ < PCelt. **arto-* < **arxto-* < **h₂rtkō-*. This reduction explains the development of the *t*-preterite to roots ending in liquids, departing from the original 3sg. *s*-aorist, **b^hēr-s-t* > **bīr-t* → **bir-t-* (OIr. *birt*, *-bert*, MW *kymyrth*).

Inherited **st* appears to have been preserved as such in Proto-Celtic, as indicated by its survival in e.g. Celtib. *boustom* ‘cow stable (?)’ < PCelt. **bousto-* < PIE **g^wou-s^hth₂-o-* (Ved. *goṣṭhá-*) and occasional survival in Brittonic. In Goidelic, on the other hand, **st* has given **ss* > *s* in all positions. PIE **-Dt-*, presumably realized as [tst] and hence indistinguishable from **-Dst-*, was reduced to PCelt. **-ts-*. In Insular Celtic, if not earlier, this was further reduced to **ss*, as in PIE **uid-to-* > **uisso-* > OIr. *-fess* ‘is known’, MW *gwys*, MBret. *gous* ‘was known’.

9.2.8 Elimination of PIE *p

The elimination of the phoneme *p is probably the best-known defining sound change in the Celtic languages. However, in many contexts *p leaves a trace by merging with other phonemes, e.g. PIE *Vpn > PCelt. *V_{un} (PIE *s_uop-no- or *sup-no- > PCelt. *so_uno- > OIr. *súan*, MW *hun* ‘sleep’),² *VpL > *VbL (PIE *d_{uej}-plo- > *d_{uej}blo- > OIr. *diabul* ‘twofold’), *ps > *xs (cf. PCelt. *o_{ux}selo-above), *pt > *xt (cf. PCelt. *s_{ext}am above). It is possible that the phoneme /p/ was not completely lost by Proto-Celtic times but that we still had /p/ in the earliest attested stage. This is primarily based on Lep. *uvamokozis*, plausibly analysed as a personal name with a first member /u_φamo-/ from PIE *up-m_{Ho}- ‘highest’ (cf. Schumacher 2004: 133–4; Eska 2013). Another indication that a reflex of PIE *p was still preserved as a distinct phoneme in Proto-Celtic is provided by the reflex of PIE *sp. In initial position, this cluster yields OIr. s-, len. f- and Brit. *f-, as in OIr. *seir* ‘heel’, dual *dī p[h]erith*, W *ffēr* ‘ankle’ < PCelt. *s_{ph}eret- < PIE *sp_{her}H- ‘to kick’ (LIV² 585), OIr. *selg*, MBret. *felch* ‘the spleen’ < PCelt. *s_{ph}elgā < PIE *spelg^h- (cf. Lat. *lien*, Ved. *plīhān*-, Gr. *σπλήν*). This distribution of outcomes does not correspond with that of any other known initial cluster. We cannot posit PCelt. *s_u- as the outcome of PIE *sp- because, while PCelt. *s_u- would account for Old Irish s-, len. f- (as in OIr. *siur*, len. *do phethar* < PIE *s_{ues}or-), it gives *h_u- in Brittonic (MW *chwaer*, MBret. *hoar* ‘sister’) and cannot therefore have merged with the outcome of PIE *sp-. Indeed, PIE *sp- appears to be the only regular source of PBrit. *fV- apart from Latin borrowings with f-.

9.2.9 Length Opposition in Consonants

A length opposition in sonants had already developed in pre-Proto-Celtic. The long sonants came about by assimilation, the most common being the assimilation of *-sR- to *-RR-, as in Celtib. *iomui* < PIE *i_{osm}ōi-, OIr. *coll*, OW *coll*, OBret. *coll-guid* ‘hazel-tree’ < *kollo- < pre-PCelt. *koslo-. Hence, Proto-Celtic acquired an opposition between *n* and *nn*, *m* and *mm*, *l* and *ll*, and *r* and *rr*. PIE postvocalic *-sr- may, however, have yielded *-dr- instead, as indicated by Gaul. *tidres*, OIr. *teoir*, MW *teir* fem. ‘three’ < *tidres < PIE *tisres (Schrijver 1995: 448–52). However, *rr developed from other sources, such as PIE *rs (OIr. *carr*, MW *karr*, MBret. *carr* ‘cart’, Latin *carrus* from Gaulish < PCelt. *karro- < PIE *krso-/k_ṛso-) and possibly *rp.

The phonemic length opposition in sonants is paralleled by a similar opposition in stops. Proto-Indo-European did not allow geminate stops, at least not outside Lallwörter, but new geminate stops arose at some point in Celtic. This

² Probably only after rounded vowels, cf. PIE (Transponat) *tep-net- > PCelt. *tenet- > OIr. *teine* ‘fire’).

happened either by regressive assimilation between two stops across a morpheme boundary, e.g. PIE (Transponat) $*(h_2)ad-\hat{k}(i)jah_2$ ‘at-ness’, ‘proximity’ > $*akkjā$ > OIr. *aicce* ‘proximity; fosterage’, MW *ach* ‘lineage, ancestry’ or through hypocoristic gemination observable particularly in personal names.

9.2.10 *Lenition of Voiced Stops*

A purely phonetic lenition of the short voiced stops after vowels may possibly be reconstructed for Proto-Celtic or a Common Celtic period shortly thereafter.³ While this lenition is central to Insular Celtic, operating both word-internally and across word boundaries as part of grammatical lenition, there is some evidence in favour of it going back much further. The use of an apparent sibilant symbol for the outcome of mainly postvocalic $*d$ in Celtiberian (Villar 1995) may plausibly be interpreted as an indication of phonetic lenition to [ð]. Likewise, the occasional loss of intervocalic $*g$ (as in Celtib. *tuateres* ‘daughters’ < PCelt. **dugateres*) may be an indication of intervocalic /g/ being realized as [ɣ]. It is likely that this lenition also affected $*s$ (> $*h$) and $*m$ (> $*\beta$), as it did in Insular Celtic. The occasional loss of /s/ in Gaulish may support this.

9.2.11 *Morphological Innovations*

As noted above, in many instances where Brittonic and Goidelic share morphological innovations, it is difficult to tell whether or not these innovations date back to Proto-Celtic, Common Celtic or a later stage common to Goidelic and Brittonic. The following non-trivial innovations are likely to date back to Proto-Celtic or at least an early Common Celtic stage:

- Levelling of the pronominal $*so-/to-$ paradigm in favour of the allomorph with $*s-$, as evident from e.g. Celtib. dat.sg. *somui* < PCelt. $*sommūi$ ← PIE $*tosmōi$, the OIr. 3pl. prepositional dative ending *-ib* and the PBrit. 3pl. prepositional ending $*-u\beta$ (MW *-udd*, MBret. *-e*, *-o*, *-eu*)⁴ < PCelt. dat.pl. $*soibis$ ← PIE $*toibʰi(s)$. This innovation is possibly shared with Italic.
- Loss of the agent noun suffix $*-ter-/tel-$: while there are many instances of the instrument-noun suffix $*-tro-/tlo-$ in Celtic, the agent-noun suffix $*-ter-/tel-$, from which $*-tro-/tlo-$ is plausibly derived, is completely absent from Celtic. Instead, we find alternative productive suffixes with this function, such as $*-ijo-$ and $*-ijati-$ (abstracted from $*CeCH-ti-$ > $*CeC-ati-$ and suffixed to $*-ijo-$ stems).

³ The Common Celtic period refers to the period after the split-up of Proto-Celtic in which innovations could still affect all Celtic varieties.

⁴ Cf. the treatment of final unstressed $*-lu\beta$ ‘herb’ in $*tud-lu\beta$ ‘navelwort’ > ModBret. *tule*, *tulo*, *tulev*.

- Elimination of the present and past active participle as part of the verbal paradigm. The former survives in numerous fossilized nominal formations, e.g. PCelt. **karant-* ‘friend’ (OIr. *carae*, W *car*). The past passive participle is preserved in this function, though typically in the form **-t̃io-* for expected **-to-* in Insular Celtic.
- Merger of the aorist and the perfect into a preterite (cf. the parallel innovations in the prehistory of Italic and Germanic).
- Loss of the inherited categories of subjunctive and optative (although the Celtic *s-* and *ā-*subjunctives and futures may continue PIE *s-*aorist subjunctives).
- *s-*aor.3sg. **-s-t* reanalyzed as **-st-* and used as a marker for the past tense.
- Thematization of **es-* ‘is’: there is evidence from both Old Irish and Brittonic that at least some of the present-tense forms of **es-* (PIE **h₁es-*) were thematized to **es-e/o-*, as described by Schrijver 2020.

9.3 The Internal Structure of Celtic

The precise internal subgrouping of Celtic is still not entirely settled (cf. the tentative tree in Figure 9.1). However, it seems fairly clear that Celtiberian should be contrasted with the more northern varieties (cf. Schrijver 2015; Eska 2017). This may be demonstrated for instance by the development of a clitic relative particle **iō* in Gaulish, Goidelic and Brittonic as opposed to the fully inflected relative pronoun **iō-* in Celtiberian (e.g. dat.sg. *iomui*) and the transfer of the feminine gen.sg. ending **iās* from the *ī-*stems to the *ā-*stems (Gaul. *-ias*, OIr. *-e* as opposed to Celtib. *-as*). Celtiberian, conversely, has innovated e.g. by creating a new *o-*stem gen.sg. in *-o* of unclear origin.

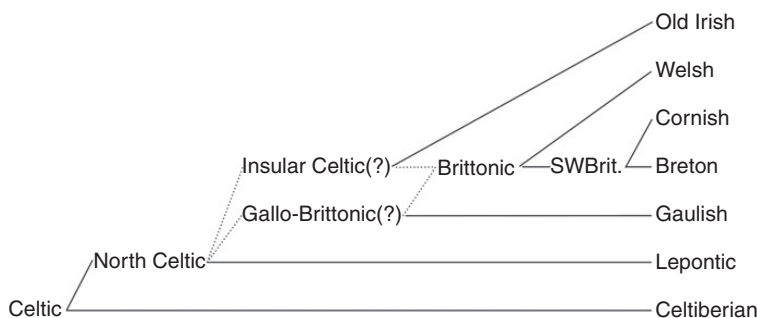


Figure 9.1 The Celtic languages

9.3.1 *Goidelic*

Establishing Goidelic as a branch separate from the other Celtic branches is unproblematic, Goidelic being easily defined by a long series of sound changes differentiating it from Proto-Celtic and resulting in the remarkably uniform Old Irish language. Recent chronological overviews of these developments are given in [Sims-Williams \(2003: 296–301\)](#) and [Stifter \(2017: 1198–200\)](#). The subsequent Goidelic dialects derive more or less directly from Old Irish. Among the defining features of Old Irish, we can include the following, in rough chronological order (largely following [McCone 1996: 105–25](#); but cf. e.g. [Kortlandt 1997](#) and [Isaac 2007: 97–113](#)):

- Fronting and raising of PCelt. **a* to **æ* before tautosyllabic nasals. This **æ* may subsequently be further raised to *e/i* or *é* (when lengthened) or revert back to *a*, the conditions for this being debated (cf. [McCone 1992](#); [Schrijver 1993](#)). Word final **-an* (as in the nom.-acc.sg. of neuter *n*-stems and the acc. sg. of *ā*-stems and consonant stems) is also affected by this, giving **-en* which usually causes palatalization when lost by apocope.
- **o* > **a* in final syllables.
- *VNT* > *V(·)D*, i.e. loss of nasals before voiceless obstruents (PCelt. **k, *kʷ, *t, *s, *x*), with voicing of a following stop and frequently with compensatory lengthening of a preceding front vowel, e.g. PCelt. **kanto-* ‘100’ > **kento-* > OIr. *cét* /kʲeːd/, PCelt. **krenxtV-* > OIr. *crécht* ‘wound, scar’ (MW *creith*, MBret. *creizenn*).
- Postvocalic lenition of voiceless stops to the corresponding voiceless fricatives.
- Raising and lowering of short vowels caused by the height of the vowel in the following syllable.
- Several rounds of palatalization, whereby consonants are palatalized by front vowels under different circumstances. The front vowels may subsequently be lost (by syncope or apocope) or reduced to schwa, causing the palatalization to become phonemic.
- Loss of the rounding of the reflexes of PCelt. **kʷ, *gʷ* and merger with the plain velars. In some cases, the rounding may be transferred to a following vowel, as in PCelt. **kʷrimi-* (MW *pryf*) > OIr. *cruim* /kruɲ/ ‘worm’.
- Apocope of vowels in absolute auslaut. Long vowels followed by a consonant are shortened but preserved.
- Loss of fricatives before sonants with compensatory lengthening or *u*-diphthongization of the preceding vowel, e.g. **(ϕ)etnos* > **eθnah* > OIr. *én* ‘bird’ (MW *edyn*, MBret. *ezn*), gen.sg. **(ϕ)etnī* > **eθnī* > *euin*. PCelt. **tr* > *θr* is not affected by this change, cf. PCelt. **aratrom* (MW *aradyr*, MBret. *arazr*) > **araθran* > OIr. *arathar* ‘plough’.

- Syncope of vowels in even-numbered, medial syllables, after the operation of apocope.
- Initial, unlenited **u-* > OIr. *f-* (presumably /*ϕ-*/), PCelt. **uiro-* > OIr. *fer*.

9.3.2 Brittonic

The existence of a Brittonic subgroup distinct from Goidelic is uncontroversial, even if the position of Brittonic is itself contested. Much like Goidelic, Brittonic underwent a fundamental transformation in the early Medieval period. Unlike Old Irish, however, where the inherited nominal and verbal morphology remained largely intact, albeit in a much-altered guise, the sound changes in Brittonic resulted in massive restructurings in inflectional morphology, leading to a complete loss of the nominal case system. The singular/plural opposition has also been partially restructured: in some nouns, typically ones that would often occur in the plural, the undervived form has plural (or “collective”) meaning, with the singular (or “singulative”) being formed by a Proto-Brittonic suffix *-*inn* (masc.), *-*enn* (fem.), as in e.g. coll. **guið* ‘trees’, sglt. **guið-enn* ‘a tree’ (W *gwydd*, *gwydden*, MBret. *guez*, *guezenn*) < PCelt. **uidu-* (OIr. *fid*).

These defining sound changes took place after the introduction of the main body of Latin loanwords, since these loanwords are generally affected by the same changes as inherited vocabulary. Many of the changes may be due to contact with early Gallo-Romance, such as voicing of postvocalic voiceless stops, penultimate stress, loss of phonemic vowel length and the loss of the neuter gender, cf. Schrijver 2002. The phonological changes leading from Proto-Celtic to Old Welsh, Old Cornish and Old Breton have been treated in great detail by Jackson 1953: 265–699, Schrijver 1995 and Sims-Williams 2003. Chronological overviews are given in Jackson 1953: 694–99 and Stifter 2017: 1200–1.

- The Proto-Celtic voiced geminate stops appear to have been devoiced in Brittonic (cf. Pedersen 1909: 159–61) and subsequently fricativized regularly, just like the Proto-Celtic voiceless geminates (cf. spirantization below). This is borne out by e.g. PCelt. **biggo-* (OIr. *bec* /b’eg/) > **bikko-* > PBrit. **bix-an* (dimin. suff. *-*an*; W *bychan*, Bret. *bihan* ‘little’), PCelt. **kloggā* (OIr. *cloc*, *clog* /klog/) > **klokkā* > PBrit. **klox* (W *cloch*, Bret. *kloc* ‘h’ ‘bell’), PCelt. **uraggā* (Ir. *frac*, *frag* /frag/) > **urakkā* > PBrit. **urax* (W *gwrach*, Bret. *gwrac* ‘h’ ‘hag’), PCelt. **buggo-* (OIr. *bog* ‘gentle, tender’) > **bukko-* > Brit. **bux* (ModBret. *bouc* ‘h’ ‘blunt’; Bret. *bouk* ‘soft’ must instead be a borrowing from Irish). This change also accounts for the development of PCelt. **zd* (**ǵdʰ*) in Brittonic, which appears to have gone through **dd* (OIr. /d/) to **tt* and ultimately to PBrit. **θ*. This may be exemplified by e.g. PCelt. **nizdo-* > **niddo-* (OMIr. *net*, *ned* /N’ed/) > **nitto-* > PBrit. **nιθ* (W *nyth*,

Bret. *neizh* ‘nest’). The change must predate the creation of geminate stops by assimilation of preverbs and verbs and the univerbation of verbal compounds such as PCelt. **kred-dī-* ‘believes’ (MW *credu*, MBret. *cridiff*, OIr. *creitem* ‘to believe’).

- Restructuring of the vowel system: fronting of PCelt. **ū* > **ī* (causing a merger with PCelt. **ī*) and monophthongization of the remaining diphthongs: PCelt. **oi* > **ō* (merging with the reflex of PCelt. **ou*) > **ū*, PCelt. **ai* > **ē* and PCelt. **au* > **ɔ̄* (possibly via **ā*).
- Final *a*-affection: a Proto-Celtic long **ā* in the final syllable lowers a preceding PCelt. **i*, **u* to **e*, **o*. This may be observed in feminine *ā*-stems, especially in adjectives in Middle Welsh, where the lowering has become a mark of the feminine, e.g. nom.sg.masc. **ūind-os*, nom.sg.fem. **ūind-ā* > MW masc. *gwynn*, fem. *gwenn* ‘white’.
- Lenition of postvocalic voiceless stops to the corresponding voiced ones, e.g. PCelt. **brātīr* > PBrit. **brɔdr* ‘brother’ (MW *brawd*, MBret. *breuzr*), PCelt. **dekam* > PBrit. **deg* ‘ten’ (MW *dec*, MBret. *dec* /deg/).
- Nasalization of voiced stops, *ND* > *NN*, as in PCelt. **landā* (Mlr. *land*, Gaul. **landā* ⇒ Fr. *lande* ‘heath’) > MW *llan* ‘enclosure; church’, MBret. *lann*. This also operates in syntactically close external sandhi and gives rise to the limited Brittonic nasal mutation.
- Fixed stress on the penultimate syllable. With apocope (see below), the stress came to fall on the final syllable.
- Final *i*-affection, whereby a short vowel is raised and/or fronted by a final **-ī* and **-io-*. After apocope, a new round of *i*-affection takes places, this time caused by high front vowels still remaining after apocope.
- Apocope of all final syllables. In contrast to Goidelic, even long vowels followed by consonants are lost.
- Syncope of immediately pretonic vowels in open syllables.
- Spirantization or “second lenition”, whereby previously unlenited voiceless stops become voiceless fricatives after vowels and non-homorganic fricatives (thus Schrijver 1999; for a different chronology, see Sims-Williams 2007: 43–58). This includes former geminate stops and stops protected from the first lenition in external sandhi. It is possible that this development was sufficiently late to have developed differently in Welsh and South-West Brittonic: in external sandhi, spirantization only seems to occur after vowels in Welsh, while in South-West Brittonic it also appears to take place after non-homorganic sonants.
- The new quantity system, whereby the inherited phonemic vowel length was lost. However, this did not entail any large-scale merger, since older phonemic contrasts in length were shifted to quality (e.g. PCelt. **ī* > **i* as opposed to PCelt. **i* > **ɪ*, PCelt. **ū* > **u* as opposed to PCelt. **u* > **u* and PCelt. **ā/au* > **ɔ̄* > **ɔ* as opposed to PCelt. **o* > **o*) or preserved by diphthongization of

the old long vowels (PCelt. $*e_i > *ē >$ PBrit. $*u_i$ and PCelt. $*a_i > *ē > *o_i$). The new quantity system only has allophonic vowel length, with stressed vowels being long before single consonants and in word-final position and short elsewhere.

A number of changes take place in the course of the Old British transmission but are nevertheless shared by all Brittonic branches, e.g.:

- Initial, non-lenited $*u- > *gu-$, as in PCelt. $*u_{i}ro- >$ PBrit. $*u_{i}ur >$ MW *gwr* ‘man’.
- Accent retraction from the final to the penultimate syllable. It is unclear whether the final stress of Vannetais Breton is the result of a later forwards shift due to French influence or if it represents an archaism, with the Proto-Brittonic final stress being preserved due to a higher proportion of French speakers in this region.

Though it may seem surprising at first glance, given the geographical proximity of Cornwall to Wales and its relatively long distance from Brittany, there is a fair amount of evidence in favour of a distinct South-West Brittonic branch consisting of Cornish and Breton to the exclusion of Welsh (cf. [Hamp 1953](#), [Jackson 1953](#): 19–25 and *passim*, [Schrijver 2011](#): 15–33).

9.3.3 The Position of Brittonic: Gallo-Brittonic or Insular Celtic?

The position of Brittonic in the Celtic family tree remains an unsolved question, specifically whether we should posit an Insular Celtic node consisting of Brittonic and Goidelic to the exclusion of Gaulish (as e.g. [McCone 1992](#); [Schrijver 1995](#): 463–5; [Eska 2017](#)), a Gallo-Brittonic node excluding Goidelic (as [Koch 1992](#)) or a dialect continuum with a fundamental three-way split, allowing Brittonic to share innovations with both Gaulish and Goidelic (thus e.g. [Sims-Williams 2007](#): 34).

Among the potential Gallo-Brittonic innovations are the following:

- $*k^w > *p$ in Gaulish, Leponic and Brittonic, as in PCelt. $*ek^wo-$ ‘horse’ > Gaul. *Epona* ‘name of a goddess’, MW *ebawl* ‘foal’ (cf. OIr. *ech* ‘horse’).
- A change of $*oRa$ to $*aRa$, i.e. an expanded Joseph’s Law, seems to occur in Brittonic and Gaulish, as shown by MW *taran* ‘thunder’, Gaul. *Taranis* as opposed to OIr. *torann* ‘thunder’.
- PCelt. $*sr- > *fr-$ seen in e.g. PCelt. $*sroKnā$ (OIr. *srón* ‘nostril’) > Brit. $*from$ (W *ffroen*, MBret. *froan* ‘nostril’), Gaul. $*frognā$ (whence OFr. *frongne* ‘scowl, frown’).
- Devoicing of the voiced geminate stops: we may assume that this change also took place in Gaulish, thus providing us with a potential Gallo-Brittonic isogloss. This is based on the evidence of PCelt. $*k^wezdi-$ ‘bit, piece’ > $*k^weddi-$ (OIr. *cuit* /kud/) > Gallo-Brit. $*pettiā$ (Gaulish $\Rightarrow$ LLat. $*pēttiā >$

Fr. *pièce*, etc.), Brit. **peθ* (W *peth*, Bret. *pezh*) and PCelt. **bozdo-* ‘knob’ > **boddo-* (Mlr. *bot* /bod/ ‘tail; membrum virile’) > Gallo-Brit. **botto-* (Gaulish ⇒ LLat. **bottu-* > Fr. dial. *bo*, *bout* ‘hub of a wheel’, *bouton* ‘button’), Brit. **boθ* (W *both* ‘hub of a wheel’), cf. [Delamarre 2003](#): 93, 249–50. Even if one does not accept a general devoicing of voiced geminates in Gallo-Brittonic, the specific development of PCelt. **zd* to **tt* still constitutes an isogloss.

- Thematisation of feminine consonant stems: a few consonant stems, preserved as such in Goidelic, seem to have been transferred to the feminine *-ā*-stems in Brittonic and Continental Celtic. The trigger for the transfer was probably the Proto-Celtic acc.sg. **-am* (< PIE **-m*) and acc.pl. **-ās* (< PIE **-ms*), which had become identical to the feminine *-ā*-stems. Examples include PCelt. **abū*, **abon-am* (OIr. *aub*, *abainn*) → **abon-ā* (MW *afon*, MBret. *auon* ‘river’, Gaul. **abonā* ⇒ Fr. *Avosnes*, name of a village), PCelt. **brix-s*, **brig-am* (OIr. *brí* ‘hill’) → **brig-ā* (MW *bre*, MBret. *bre* ‘hill’, Gaul. *-briga* in numerous place names) and PCelt. **brus-ū*, **brunn-am* (OIr. *brú*, *broinn*⁵ ‘belly; womb’) → **brunnā* (MW *bron*, MBret. *bronn* ‘breast’, Gaul. **brunnā*, possibly reflected in Modern Gallo-Romance, cf. [von Wartburg 1928](#): 566).

The list of potential shared innovations between Gaulish and Brittonic may not be particularly impressive, yet it should be noted that it is difficult to point to any significant Gaulish innovations *not* shared with Brittonic. One might even pose the question as to whether Brittonic could simply be seen as continuing a dialect of Gaulish. Such a scenario would require the following Insular Celtic innovations to have resulted from a later Sprachbund-type situation:

- The absolute/conjunct opposition, whereby many finite verbal forms have longer endings when in initial position of the verbal syntagm. This is in all likelihood an innovation of “Insular Celtic”, brought about by the generalization of a main clause verbal particle **et(i)*, which occupied the second position of the clause, protecting the verbal endings from reduction when the verb is clause initial ([Schrijver 1994](#); 1997: 147–58; [Schumacher 1999](#); 2004: 90–114).
- Striking similarities in the system of verbal morphology, particularly with regard to the formation of compound verbs, perfective particles and infix pronouns. There is little to no evidence for this from Continental Celtic.

⁵ It is conceivable that the Old Irish paradigm Nsg. *brú*, Gsg. *bronn* reflects a remodelled PCelt. **brunn-s*, **brunn-os* with the Nsg. levelled after the oblique cases, rather than the spectacularly archaic and irregular PCelt. **brus-ū*, **brun-n-os*. Irrespective of this, the ultimate origin of either Proto-Celtic paradigm must be a PIE (Transponat) **bʰrus-ō*, **bʰrus-n-os*.

- Analogical replacement of **uer* ‘over’ (PIE **uper*) with **uor* (OIr. *for*, MW *gor-*, *gwar-* as opposed to Gaul. *uer-*), probably under the influence of the antonym **uo* ‘under’ (PIE **upo*).

9.4 The Relationship of Celtic to the Other Branches

With the exception of Italic (cf. [Chapter 7](#)), no branch of Indo-European appears to share a significant number of isoglosses with Celtic, at least not to the extent that a plausible case can be made for a shared post-Proto-Indo-European stage.

Admittedly, there does appear to be a special connection between Celtic and Germanic to the exclusion of other branches. However, the shared features are almost exclusively lexical in nature (for Dybo’s Shortening see [Section 9.2.3](#)), either the existence of a root, e.g. **tegu-* ‘fat, thick’ (OIr. *tiug*, W *tew*; OE *þicce*, OHG *dicki*), **magu-* ‘boy, servant’ (OIr. *mug*, MCorn. *maw*; Goth. *magus*, OE *magu*),⁶ or a specific semantic development encountered only in these two branches, such as **priH-o-* ‘beloved’ (Ved. *prīyá-*) > ‘free’ (W *rhydd*, Goth. *freis*, OHG *frī*). The absence of any securely identified innovations in the realm of inflectional morphology between Celtic and Germanic makes it very likely that this relatively impressive collection of lexical isoglosses is due to borrowing.

Apart from lexical isoglosses, there are a few apparent shared innovations which are worth mentioning:

- A notable syntactic correspondence between Hittite and Old Irish is the use of PIE **nu* (Hitt. *nu*, OIr. *no*) as a sentence initial particle. In Old Irish, this is done in order to provide a preverb to which a clitic pronoun can be attached, while in Hittite it functions as a sentence connecting particle to which clitics may be suffixed.
- Another notable syntactic correspondence is between Celtic and Tocharian. This is the development of the PIE adverb **(h₁)eti* (Lat. *et*, Goth. *ip*, Ved. *áti*) to a clitic obeying Wackernagel’s Law. In Insular Celtic this yields the main clause particle **et(i)*, blocking lenition of the following element, responsible for the emergence of the absolute/conjunct allomorphy in the finite verb in Insular Celtic, in Tocharian B it produces *-s* ‘and’, a clitic connector following the first word of the clause ([Hackstein 2005](#): 176).

9.5 The Position of Celtic

See [Chapter 7](#).

⁶ One could consider **magu-* to be a late borrowing from Germanic to Common Celtic. This would allow us to reconstruct pre-Germ. **mh₂k-u-* ‘a reared one’ > PGerm. **magu-*, from the PIE root **mah₂k-* ‘to rear, to nourish’. The inherited Celtic cognate would be PCelt. **mak^wo-* ‘boy, son’, continuing a thematized **mh₂k-u-*.

References

- de Bernardo Stempel, Patrizia. 1987. *Die Vertretung der indogermanischen liquiden und nasalen Sonanten im Keltischen*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Delamarre, Xavier. 2003. *Dictionnaire de la langue gauloise*. Paris: Errance.
- Dybo, V. A. 1961. Sokrašćenie dolgot v kel'to-italijskix jazykax i ego značenie dlja balto-slavjanskoj i indoevropskoj akcentologii [The shortening of long vowels in Italo-Celtic and its significance for Balto-Slavic and Indo-European accentology]. *Voprosy slavjanskogo jazykoznanija* 1961(5). 9–34.
- Eska, Joseph F. 2013. In defense of Celtic /φ/. In Adam I. Cooper, Jeremy Rau & Michael Weiss (eds.), *Multi nominis grammaticus: Studies in Classical and Indo-European linguistics in honor of Alan J. Nussbaum on the occasion of his sixty-fifth birthday*, 32–43. Ann Arbor, MI: Beech Stave.
- Eska, Joseph F. 2017. The dialectology of Celtic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1264–74. Berlin: De Gruyter Mouton.
- Hackstein, Olav. 2005. Archaismus oder historischer Sprachkontakt – Zur Frage west-indogermanisch-tocharischer Konvergenzen. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 169–84. Wiesbaden: Reichert.
- Hamp, Eric P. 1953. Morphological correspondences in Cornish and Breton. *Journal of Celtic Studies* 2. 5–24.
- Hill, Eugen. 2012. Silbische Liquiden vor Nasalen im Inselkeltischen und das Problem der Nasalpräsentien vom Typ air. *sernaid*, kymr. *-sarnu*. *Keltische Forschungen* 5. 157–84.
- Isaac, Graham R. 2007. *Studies in Celtic sound changes and their chronology*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Jackson, Kenneth. 1953. *Language and history in early Britain*. Edinburgh: Edinburgh University Press.
- Joseph, Lionel S. 1982. The treatment of *CRH- and the origin of CaRa- in Celtic. *Ériu* 33. 31–57.
- Koch, John T. 1992. “Gallo-Brittonic” vs. “Insular Celtic”: The inter-relationships of the Celtic languages reconsidered. In Gw. Le Menn & J.-Y. Le Moing (eds.), *Bretagne et pays celtiques – langues, histoire, civilisation: Mélanges offerts à la mémoire de Léon Fleuriot*, 471–95. Saint-Brieuc & Rennes: SKOL & Presses universitaires Rennes.
- Kortlandt, Frederik. 1997. On the relative chronology of Celtic sound changes. *Historische Sprachforschung* 110. 248–51.
- LIV² = Helmut Rix et al. (eds.). 2001. *Lexikon der indogermanischen Verben*. 2nd ed. Wiesbaden: Reichert.
- McCone, Kim. 1992. Relative Chronologie: Keltisch. In Robert S. P. Beekes, Alexander M. Lubotsky & Jos Weitenberg (eds.), *Rekonstruktion und relative Chronologie*, 11–39. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- McCone, Kim. 1996. *Towards a relative chronology of ancient and medieval Celtic sound change*. Maynooth: Department of Old Irish, St. Patrick's College.
- Pedersen, Holger. 1909. *Vergleichende Grammatik der keltischen Sprachen*. Vol. 1. Göttingen: Vandenhoeck & Ruprecht.
- Schrijver, Peter. 1993. On the development of vowels before tautosyllabic nasals in primitive Irish. *Ériu* 44. 33–52.

- Schrijver, Peter. 1994. The Celtic adverbs for 'against' and 'with' and the early apocope of *-i. *Ériu* 45. 151–89.
- Schrijver, Peter. 1995. *Studies in British Celtic historical phonology*. Amsterdam: Rodopi.
- Schrijver, Peter. 1997. *Studies in the history of Celtic pronouns and particles*. Maynooth, Dept. of Old Irish, National University of Ireland.
- Schrijver, Peter. 1999. Spirantization and nasalization in British. *Studia Celtica* 33. 1–19.
- Schrijver, Peter. 2002. The rise and fall of British Latin. In Markku Filppula, Juhani Klemola & Heli Pitkanen (eds.), *The Celtic Roots of English*. Joensuu: University of Joensuu.
- Schrijver, Peter. 2011. Old British. In Elmar Ternes (ed.), *Brythonic Celtic – Britannisches Keltisch*, 1–84. Bremen: Hempen.
- Schrijver, Peter. 2015. Pruners and trainers of the Celtic family tree: The rise and development of Celtic in the light of language contact. In Liam Breatnach et al. (eds.), *Proceedings of the XIV International Congress of Celtic Studies*, 191–219. Dublin: Dublin Institute of Advanced Studies.
- Schrijver, Peter. 2020. Italo-Celtic and the inflection of *es- 'be'. In Matilde Serangeli & Thomas Olander (eds.), *Dispersals and diversification: Linguistic and archaeological perspectives on the early stages of Indo-European*, 209–35. Leiden: Brill.
- Schumacher, Stefan. 1999. Randbemerkungen zu absolut und konjunkt: Mittelmymrisch *hanfot*. In Peter Anreiter & Erzsébet Jerem (eds.), *Studia Celtica et Indogermanica: Festschrift für Wolfgang Meid zum 70. Geburtstag*, 453–64. Budapest: Archaeolingua.
- Schumacher, Stefan. 2004. *Die keltischen Primärverben*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Sims-Williams, Patrick. 2003. *The Celtic inscriptions of Britain: Phonology and chronology, c 400–1200*. Oxford: Blackwell.
- Sims-Williams, Patrick. 2007. *Studies on the Celtic languages before the year 1000*. Aberystwyth: CMCS.
- Stifter, David. 2017. The phonology of Celtic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1188–202. Berlin: De Gruyter Mouton.
- Villar, Francisco. 1995. *A new interpretation of Celtiberian grammar*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- von Wartburg, Walther. 1928. *Französisches etymologisches Wörterbuch: eine Darstellung des galloromanischen Sprachschatzes*. Vol. 1. A–B. Basel: Zbinden.
- Zair, Nicholas. 2012. *The reflexes of the Proto-Indo-European laryngeals in Celtic*. Leiden: Brill.

10 Germanic

*Bjarne Simmelkjær Sandgaard Hansen &
Guus Jan Kroonen*

10.1 Introduction

Germanic languages are spoken by about 500 million native speakers. They constitute a medium-large subgroup of the Indo-European language family and were originally located in Northern Europe, owing much of their current distribution to the recent expansion of English. From a historical perspective, notable old Germanic languages were Gothic, Old Norse, Old English, Old Frisian, Old Saxon, Old Franconian (poorly attested) and Old High German (Bousquette & Salmons 2017: 387–8). Gothic, mainly known from a fourth-century Bible translation, continued to be spoken in a local variant in Crimea until the late eighteenth century but subsequently went extinct (Nielsen 1981: 283–8). The remaining old Germanic languages developed into modern varieties such as English, Frisian, Dutch, German and the Nordic languages (Henriksen & van der Auwera 1994).

However, Northern Europe must have witnessed speakers of some form of Germanic even prior to the attestation of these old Germanic languages. Runic inscriptions in a language that we may label Early Runic appear from the second century onwards, and one inscription on a fourth-century-BCE bronze helmet, the Negau B helmet, has been unearthed in Slovenia. This inscription, which is in a northern Etruscan alphabet and reads **hariḡastiteiwa**, constitutes our earliest evidence of Germanic, at least if we follow Markey (2001) in interpreting it as ‘Harigast the priest’.¹ It thus constitutes a *terminus ante quem* for some of the linguistic features that define Germanic (Section 10.2).

10.2 Evidence for the Germanic Branch

In this section, we shall list some of the most important diagnostic features of Germanic within the realms of phonology and morphosyntax, which constitute the most reliable means for establishing linguistic clades (see Section 2.3).

¹ With *harigast* as the Germanic words for ‘army’ and ‘guest’ and *teiwa* as a linguistic precursor of the Nordic theonym *Týr*. Alternatively, Must (1957: 55–7) sees a Rhaetic name consisting of Venetic and Etruscan elements in this inscription.

10.2.1 Phonology

As indicated by Ringe (2017: 113–27, 147–50 and Section 4.3), all Germanic languages display reflexes of the outputs of the following phonological innovations.² Three of these, no. 1, 4 and 5, are frequently said to constitute the most striking hallmark of the Germanic languages, i.e., “what to a large extent defines Germanic” (Kroonen 2013: xxvii).

1. Rask/Grimm’s Law I: PIE $*p\ t\ k/k\ k^w >$ fricatives $*f\ p\ h\ hw$ unless an obstruent immediately preceded, e.g. Goth. *fadar* ‘father’ ~ Ved. *pitā*, Gr. *πατήρ*; Goth. *þreis* ‘three’ ~ Ved. *tráyaḥ*, Gr. *τρεῖς*.
2. Verner’s Law: $*f\ p\ s\ h\ hw > *β\ δ\ z\ γ\ γw$ if not word-initial, if not adjacent to a voiceless sound, and if the last preceding syllable nucleus was not accented; e.g. Goth. *fadar* ‘father’ (PGmc. **faðer-* < PIE **ph₂tér-*) ~ Goth. *broþar* ‘brother’ (< PGmc. **brōþer-* < PIE **b^hréh₂ter-*).
3. Kluge’s Law: PIE **-Pn- -Tn- -Kn-* > **-bb- -dd- -gg-* (Kluge 1884; Lühr 1988; Kroonen 2011), e.g. OE *liccian*, OS *likkon*, OHG *leckōn* ‘to lick’ < PGmc. **likkōn-* < PIE **liĝh^h-neh₂-*.³
4. Rask/Grimm’s Law II: PIE $*b\ d\ g/g\ g^w > *p\ t\ k\ kw$ (including **bb\ dd\ gg > *pp\ tt\ kk*) (succeeding no. 3), e.g. Goth. *twai* ‘two’ ~ Ved. *dváu*, Gr. *δύω*; Goth. *aukan* ‘increase’ ~ Lat. *augeō* ‘increase’, Lith. *augti* ‘grow’.
5. Rask/Grimm’s Law III: PIE $*b^h\ d^h\ g^h/g^h\ g^wh >$ fricatives $*β\ δ\ γ\ γw$, e.g. OS *neþal* ‘fog’ ~ Ved. *nábhas-*, Gr. *νέφος*; Goth. *daur* ‘door’ ~ Gr. *θυρᾶ*, Lat. *forēs*.
6. $*β\ δ\ γ\ γw > *b\ d\ g\ gw$ after homorganic nasals, and $*β\ δ > *b\ d$ word-initially.
7. Shift of stress to the first syllable of the word.
8. Simplification of geminates after heavy syllables, e.g. PGmc. **wīsa-* ‘wise’ (OHG *wīs*) < **wīssa-* < PIE **ueid₂-to-*; **deupa-* ‘deep’ (Goth. *diups*) < **deuppa-* < PIE **d^heub^h-no-*.

As Ringe (Section 4.3) also mentions, the collocation of these innovations reduces the likelihood of them having taken place individually in each language – and thus the likelihood of these languages not emanating from a common predecessor – to practically zero. However, the list does not confine itself to these eight innovations. We may add at least a handful of further innovations, most of which concern the development of the inventory of stressed vowels. Examples include

² Innovations no. 3 and 8 are not mentioned by Ringe (2017). However, we have included them here to demonstrate the full range of the interdependency of these phonological innovations. The sequence of innovations no. 1–5 is disputed. Some adherents of the glottalic theory (e.g. Kortlandt 1991: 3) have Verner’s Law (no. 2) predate both Kluge’s Law (no. 3) and Rask/Grimm’s Law (no. 1, 4 and 5).

³ In view of PGmc. **seuni-* ‘sight, vision’ (Goth. *siuns*, ON *sjón*, etc.) < **sey^wni-* < PIE **sek^w-ní-*, the occurrence of Kluge’s Law must postdate innovation no. 2.

1. Merger of post-laryngeal-colouring PIE **a o ə > a*, e.g. OHG *hasō* ‘hare’ ~ Ved. *śáśa-* (< **śáśa-*), OPru. *sasins* (< PIE **kás-*); Goth. *gasts* ‘guest’ ~ Lat. *hostis* ‘enemy’ (< PIE **g^hostis*); Goth. *fadar* ‘father’ ~ Ved. *pitā́*, Gr. *πατήρ* (< PIE **ph₂tér-*).
2. Merger of post-laryngeal-colouring PIE **ā ō > *ā*.
3. **ā > *ō*,⁴ e.g. Goth. *sokjan* /sōkjan/ ‘seek’ ~ Lat. *sāgīre*; Goth. *bloma* /blōma/ ‘flower’ ~ Lat. *flōs*.
4. PIE **r̥ l̥ m̥ n̥ > *ur ul um un*, e.g. Goth. *baúrgs* /burgs/ ‘city’ ~ Av. *bərəz-* ‘high, hill, mountain’, OIr. *brí* (brig-) ‘hill’ (< PIE **b^hr̥g^h-*); Goth. *fulls* ‘full’ ~ Ved. *pūrṇáḥ*, Lith. *pilnas* (< PIE **pl̥h₁nós*).
5. Holtzmann’s Law: PIE **-j- -u- > PGmc. *-jj- -ww-* under some conditions.⁵

10.2.2 Morphosyntax

Morphosyntax, too, provides a range of compelling evidence that classifies the Germanic languages as belonging to a separate branch. A morphological innovation that may count as one of the defining hallmarks of Germanic is the rise of its verbal system. All old Germanic languages share a verbal system consisting of three subsystems:⁶

- ablauting “strong verbs” with a present stem predominantly continuing the Proto-Indo-European thematic presents and a preterite stem continuing the Proto-Indo-European perfect, e.g. Goth. *bind-an* ‘bind’, *band* ‘I/he bound’, *bund-um* ‘we bound’
- non-ablauting “weak verbs” with present stems of varying sources and preterite stems formed with a suffix containing a dental consonant, mostly in the form of reflexes of PGmc. **ð*, e.g. Goth. *haus-j-an* ‘hear’, *haus-i-da* ‘I/he heard’
- “preterite-present verbs” where the present is formed as the strong-verb preterites and the preterite as the weak-verb preterites, e.g. Goth. *kann* ‘I/he can’, *kunn-um* ‘we can’, *kun-þa* ‘I/he could’ (see also Section 10.5.2)

Although most of the building blocks of this verbal system are reflected in other Indo-European languages and thus continue Proto-Indo-European elements and processes, their regrammation and reparadigmatisation into a coherent system is

⁴ Together with a few other loanwords, Gothic *rumoneis* ‘Romans’ witnesses that innovation no. 2 is a necessary intermediary step and no. 3 must have happened after the acquaintance of the Germanic-speaking peoples with Latin. The source word, Lat. *rōmānī*, has had its *ō* rendered as *ū* (probably because innovation no. 2 caused absence of *ō* in the Germanic/pre-Gothic vowel system) and its *ā* rendered as *ō* (probably because the word was borrowed prior to innovation no. 3) (Noreen 1894: 11–12; Ringe 2017: 171; contested by Stifter 2009: 270–3).

⁵ The exact conditioning remains debated, most likely involving either adjacency to laryngeals (Hoffmann 1976: 651; Jasanoff 1978; Rasmussen 1990, 1999) or pretonic position (Kluge 1879: 128; Kroonen 2013: xxxviii–xl); see also Section 10.3.4.

⁶ In addition to these three subsystems, we find some mixed verbs and a handful of irregular verbs.

Table 10.1 *Adjectival definiteness*

Content	modifier of noun phrases (adjective) > modifier of non-definite noun phrases	individualising or characterising noun > modifier of definite noun phrases
Expression	reflexes of mainly PIE <i>a-/ō-</i> adjectival stems (= strong adj.)	reflexes of the PIE <i>n</i> -stem type (= weak adj.)

a purely Germanic innovation.⁷ So is one of the building blocks: the dental suffix found in the preterite stem of the weak verbs (Meid 1971: 107–11; Rasmussen 1996/1999; Ringe 2017: 191–4; differently Lühr 1984; Kortlandt 1989).

The system of strong and weak adjectives (Ringe 2017: 313–15) constitutes another regrammation of inherited building blocks that is highly characteristic for Germanic. Continuing mainly PIE *a-/ō-* and *n*-stems, respectively, they are not innovations formally speaking. However, the regrammation and reparaigmatisation of the function of these nominal stems (see Table 10.1) is truly innovative, as is the intrusion of pronominal endings in the strong-adjective paradigm.

Finally, degrammation and, in particular, deflection are phenomena often associated with the Germanic branch. Many of the Proto-Indo-European inflectional categories have been lost in Germanic, e.g. the aspectual system and the subjunctive mood (Ringe 2017: 177, 182–6). Others are on the verge of being lost, e.g. the mediopassive, the dual, and the vocative, ablative, locative and instrumental cases. Having arisen independently in the Germanic languages, however, these latter deflectional processes do not characterise Germanic as such. For instance, the vocative is still attested residually in Gothic, likewise the instrumental in Old High German, Old Saxon and Old English, and Early Runic may display one instance of a noun in the ablative with ablative function (Hansen 2016: 10–16). Thus, while the linguistic structures that would trigger this deflection may very well have been present in Proto-Germanic, the processes themselves occurred individually.

10.3 The Internal Structure of Germanic

It is traditionally assumed that the Germanic languages split into three sub-branches (Schleicher 1860; Streitberg 1896; Hirt 1931; etc.):

- East Germanic: the long-extinct Gothic language, Crimean Gothic and several other languages, likewise long-extinct, of which we have no or only little proof apart from toponyms and ethnonyms, e.g. Vandalic and Burgundian

⁷ For the application of the terminology of grammar, regrammation and degrammation and the connections between grammaticalisation and paradigmatisation, see Andersen 2006; Nørgård-Sørensen & Heltoft 2015.

- North Germanic: the modern-day Nordic languages Icelandic, Faroese, Norwegian, Elfdalian, Swedish and Danish and their immediate predecessors as well as the now-extinct language varieties of Norn
- West Germanic: English, Frisian (West, East and North), Dutch, Low German, High German and their various dialects, derivations and predecessors.⁸

10.3.1 *East Germanic*

Linguistic traits and developments specific for East Germanic include, within the realm of phonology, the raising of PGmc. **ǣ* to *ē* (Goth. *mena* /mēna/ ‘moon’ ~ ON *máni*, OHG *māno*), the devoicing of word-final PGmc. **-z* > *-s* (Goth. *fisks* ‘fish’ ~ ON *fiskr*, OHG *fisc*), the development of word-final PGmc. **-ō* > *-a* (neuter *a*-stem nom./acc.pl. Goth. *-a* ~ ER *-u*, ON *-Ø*⁹, OHG *-ul/-Ø*) and the absence of *a*-, *i*- and *u*-mutation (Goth. *wulfs* ‘wolf’ ~ OHG *wolf*; Goth. *gasts* ‘guest’ ~ ON *gestr*, OE *ġiest*).

Within the realm of morphology, innovations include paradigmatic levelings of the results of Verner’s Law (Section 10.2.1) in favour of the unvoiced variant (Goth. *wairþan*–*warþ*–*waúrþum*–*waúrþans* ‘become’ ~ OE *weorþan*–*wearþ*–*wurdon*–*worden*) and the creation of a deictic demonstrative pronoun Goth. *sah* ‘this’ (with *-h* < PIE **-k^we* ‘and’). We also see several instances of retention, e.g. of the reduplication in the preterite of reduplicated strong verbs (Goth. *haitan*–*haihait*–*haihaitum*–*haitans* ‘call’), of four classes of weak verbs and partially of the grammatical categories of dual and mediopassive in the inflection of verbs.

10.3.2 *North Germanic*

If we turn to North Germanic (Nielsen 2000: 255–65), some of the most salient of the many phonological innovations include loss of word-initial PGmc. **j*- (ON *ungr* ‘young’ ~ Goth. *juggs* /jungs/, OHG *jung*) and of word-initial **w*- before rounded vowels (ON *ulfr* ‘wolf’ ~ Goth. *wulfs*, OHG *wolf*), assimilation of PGmc. **-ht-* > *-tt-* (ON *nótt*, *nátt* ‘night’ ~ Goth. *nahts*, OHG *naht*), loss of word-final nasals (ON *bera* ‘carry’ ~ Goth. *baíran*), rise of *i*- and *u*-mutation with subsequent syncope or shortening of the mutation-causing unstressed vowel (PGmc. **gastiz* ‘guest’ > ON *gestr* ~ Goth. *gasts*)¹⁰ and breaking of stressed PGmc. **e* > *ja* and *jɑ* when the following syllable contained *a* and *u* prior to the aforementioned

⁸ Scholars such as Robinson (1992: 11–12), Nielsen (2000) and Bousquette & Salmons (2017: 389) express minor reservations concerning the unity of the West Germanic branch.

⁹ The superscript *u* signifies *u*-mutation on the vowels of the preceding syllable(s).

¹⁰ Similar, though not entirely identical, processes have taken place in the West Germanic languages (Section 10.3.3).

syncope, respectively (ON *jafn* ‘even, equal’ ~ Goth. *ibns* ‘even, level, flat’, OHG *eban* ‘even, equal’; ON *jörð* ‘earth, soil’ ~ Goth. *airþa* /irþa/, OHG *erda*).

On the morphological level, most of the traits that characterise North Germanic consider loss of some of the grammatical categories that were partially preserved elsewhere, e.g., the instrumental case or the dual and the mediopassive in the inflection of verbs. Others are true innovations such as the creation of a new personal pronoun in the third person (ON *hann* ‘he’, *hon* ‘she’), the replacement of the pres.3sg. ending *-þ* with pres.2.sg. *-r* > *-r* and the grammaticalisation of verbs plus reflexive pronouns into a new passive voice.

10.3.3 West Germanic

The traits and developments that justify the assumption of a West Germanic unity (Nielsen 2000: 241–7) include several innovations shared with North Germanic (Section 10.3.4). Others are not shared with North Germanic, e.g. phonological processes such as the gemination of all consonants except *r* in front of **j* (PGmc. **haffa-* ‘hold up, bear up, lift’ > OS *hebbian* ~ Goth. *haffjan*, ON *hefja*) (Krahe 1966: 95–6),¹¹ the gemination of obstruents in front of prevocalic **r* and **l* (PGmc. **bitra-* ‘sharp, bitter’ > OS *bittar*; PGmc. **apla-* ‘apple’ > OS *appul*), the rise of *i*-mutation with subsequent partial syncope or shortening of the mutation-causing unstressed vowel (PGmc. **gastiz* ‘guest’ > OE *ġiest* ~ Goth. *gasts*)¹² and the loss of word-final PGmc. **-z* in unstressed syllables prior to its merger with regular *r* (PGmc. **fiskaz* ‘fish’ > OHG *fisc* ~ Goth. *fisks*, ON *fiskr*).

The replacement of the original strong-verb pret.2sg. ending (formed by adding *-t* to the preterite singular stem) with a new one (formed by adding *-ī* to the preterite plural stem; OHG *bāri* ‘you carried’ ~ Goth. *bart*, ON *bart*; Krahe 1967: 100–3), the creation of an inflected infinitive (OHG *beranne* (dat.) ‘to bear’; Krahe 1967: 113) and the retention of reflexes of the irregular verbs PGmc. **dō-* ‘do’, PWGmc. **gā-* ‘go’ and **stā-*¹³ ‘stand’ (Krahe 1967: 137–40) constitute some of the most salient arguments from the realm of morphology.

10.3.4 Intermediary Subgroupings

It is beyond the scope of this chapter to delve into the further sub-branching of these three main sub-branches of Germanic, for which we refer to seminal works such as Nielsen (2000) instead. Rather, we shall discuss whether these three sub-branches arose simultaneously through one single ternary split or came into being through sequences of binary splits. We must therefore decide if

¹¹ North Germanic also geminates *k* and *g* in front of *j* (PGmc. **legja-* > ON *liggja* ‘lie’), but the West Germanic process applies to a much broader range of cases.

¹² Earlier in English (and Frisian) than in High and Low German (Krahe 1966: 59).

¹³ PWGmc. **stā-* ← **stō-* (< PIE **steh₂-*) by analogy with **gā-* (< PIE **ǵ^heh₁-*).

any two branches are exclusive in sharing (preferably non-trivial) phonological and morphological innovations that cannot have arisen separately in each branch.

Of the three possible combinations that could theoretically have existed, we may discard the East–West vs. North Germanic one.¹⁴ Aside from the use of the derivational nominal suffix PGmc. **-Vssu-* (Goth. *-(in)assus*, OHG *-(n)issi*), East and West Germanic share no linguistic innovations that are not also shared by North Germanic. The remaining linguistic traits shared only by East and West Germanic all constitute shared archaisms and are thus not diagnostic.

The assumption of another of the constellations, that of an initial binary split into North-East Germanic and West Germanic, once gained some popularity among Germanicists (Maurer 1942; Schwarz 1951; Rösel 1962; Lehmann 1966: 14–19; etc.) under the name Gothic–Norse theory. This split is supported by four (Schwarz 1951: 144–8) or five (Maurer 1952: 67–8) shared innovations, of which only one (Agee 2021: 337–8) may hold any diagnostic potential in a sub-branching discussion: the *Verschärfung* (i.e., occlusification) of PGmc. **-jj-* and **-ww-* > Goth. *ddj*, ON *ggj* and Goth. *ggw*, ON *ggv*, respectively, as opposed to the retention of these geminates in West Germanic where they surface as **-j-* and **-w-* (Goth. *twaddje* ‘two’ (gen.), ON *tveggja* ~ OHG *zweio*; Goth. *triggws* ‘trustworthy’, ON *tryggr* ‘trustworthy, faithful’ ~ OHG *triwi*). However, as claimed by Rasmussen (1990/1999: 383–4), the *Verschärfung* process may actually have been initiated already in Proto-Germanic, and West Germanic may have undergone a subsequent “Entschärfung” process affecting both the reflexes of PIE **-iH-* and **-uH-* and examples such as OHG *reia* ‘female roe’, OE *ræge* < PGmc. **raig-jō-*. In addition, although seemingly non-trivial, the phonological development of *Verschärfung* finds approximate parallels in Faroese (Ámason 2011: 31–3) and in the transition from Latin to Romance (Agee 2021: 338). Thus, it is if not trivial, then at least not unparalleled.

We now turn to the possibility of a North-West Germanic unity as opposed to East Germanic. More than twenty linguistic innovations appear to be shared by North and West Germanic (Agee 2021: 336). Some of these may be trivial, e.g. the lowering of PGmc. **ā* to **ā* (ON *máni* ‘moon’, OHG *māno* ~ Goth. *mena*), the development of word-final PGmc. **-ō* (via ER *-u*) > *-Ø^u* (neuter *a*-stem nom./acc.pl. ON *-Ø^u*, OHG *-Ø/-u* ~ Goth. *-a*), the rise of *a*-mutation (PGmc. **hurna-* ‘horn’ > ON *horn*, OHG *horn*)¹⁵ and perhaps even the rhotacism of

¹⁴ For a contrasting view, see Kortlandt (2000).

¹⁵ Crimean Gothic forms like *reghen* ‘rain’ and *boga* ‘arch; bow’ seem to suggest that parts of East Germanic partook in the process of *a*-mutation (Nielsen 1981: 296), thereby projecting this development back to Proto-Germanic times. The absence of short *e* and *o* in Gothic words whose North and West Germanic cognates have undergone *a*-mutation could then be due to the general Gothic merger of PGmc. **i* and **e* into **i* along with an unverifiable, but structurally expected merger of PGmc. **u* and **o* into **u*.

PGmc. *z (> r) > r (PGmc. *maizan- ‘more’ > ON *meiri*, OE *māra*) (Kümmel 2007: 80–1). On the other hand, we may not reasonably label as trivial the creation of a new deictic demonstrative pronoun by adding the enclitic particle *-si to the inherited demonstrative pronoun (Runic Danish *sasi* /sāsi/ ‘this’, OHG *dese*; Krahe 1967: 64–6) and the analogical replacement of reduplication in strong verbs by the secondary diphthong PGmc. *-ea- ~ *-ia- also known as *ē2 (ON *lét*, OHG *liaz* ‘let’ ~ Goth. *laīlot*). The latter process in particular consists of so many subprocesses that it would be inconceivable to claim independent developments in North and West Germanic. In addition, although many of the remaining shared innovations may indeed be trivial, the sheer number of instances in itself suggests a period of North–West-Germanic unity. Finally, seeing that Early Runic partakes in all the innovations common to both North and West Germanic but none of those specific to East Germanic (Nielsen 2000: 77–202, 271–98, esp. 287–93), we may safely infer that, by the time of the earliest attestations of Early Runic in the second century CE, the East Germanic branch had split off from the Germanic dialect continuum, prior to its dissolution into North and West Germanic.

On a final note, we shall review an alternative subgrouping scenario. As mentioned by Agee (2021: 344), there may still be some dialectal exchange in the years immediately following a split. If we choose to assign diagnostic value to the *Verschärfung* process, after all, and if the language varieties that would develop into the three Germanic sub-branches once coexisted in a common dialect continuum, nothing therefore prevents East and North Germanic from having shared innovations such as the *Verschärfung* at an even earlier point in time. In such a unified tree-wave model, the initial split of Proto-Germanic is defined by the first innovation (i.e., the *Verschärfung*) not shared by all its descendants, because it did not reach the entire dialect continuum. Between this initial split and the final split, which defines the exit of a dialect from the dialect continuum and thus the establishment of a separate sub-branch, the numerous innovations common to North and West, but not East Germanic, could have taken place.

Such an approach, which allows for both divergence and convergence, is also compatible with Agee’s (2021) recent glottometric calculations, which operate with degrees of subgroupiness rather than absolute, clear-cut splits. He thus (Agee 2021: 335–8, 343) posits a high subgroupiness value for North–West Germanic ($\varsigma = 20.04$) as opposed to a low value for North–East Germanic ($\varsigma = 0.032$), indicating not only that North–West Germanic is indeed a tightly knit subgroup but also that the original dialect-continuum situation may have allowed for one shared North–East Germanic innovation.

In sum, two credible models for the disintegration of Germanic present themselves. Either we must dismiss *Verschärfung* as a common North–East Germanic innovation and assume a North–West Germanic unity vis-à-vis East Germanic (as in Figure 10.1), or we must assume the existence of a Germanic

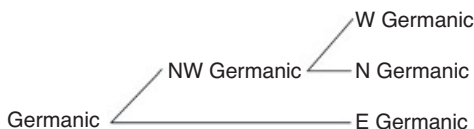


Figure 10.1 A tree model illustrating a binary split of Proto-Germanic into North-West vs. East Germanic

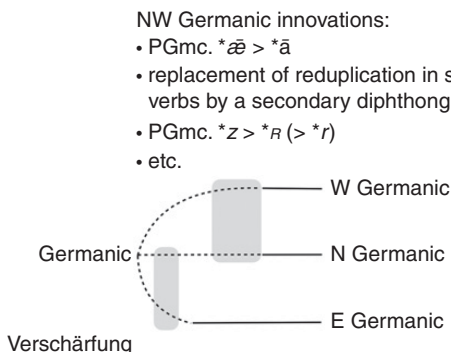


Figure 10.2 A unified tree-wave model of the Germanic dialect continuum

dialect continuum in which North Germanic could have shared innovations with first East, then West Germanic prior to the final split (as in Figure 10.2).¹⁶

10.4 The Relationship of Germanic to the Other Branches

Just as Germanic split into its sub-branches (Section 10.3), it has itself split off from Proto-Indo-European at a given point. Beyond the early divergence of Anatolian and Tocharian (Chapters 5 and 6), the relative order of the disintegration of Proto-Indo-European, including the sequence of the splits leading to Germanic, is difficult to establish, however. To solve the riddle, we must attempt to define with which other branches Germanic shares diagnostic linguistic traits, i.e., preferably non-trivial shared phonological and morphological innovations (see Section 2.3).

¹⁶ The latter model assumes that an initial split within the Germanic dialect continuum (marked by the *Verschärfung*) is followed by the numerous North-West Germanic innovations (such as PGmc. *æ > *ā and the analogical replacement of reduplication in strong verbs by a secondary diphthong) within other parts of the continuum and subsequently a final split into North, East and West Germanic.

One possibly high-node innovation that Germanic shares with several other so-called *centum* branches (Italic, Celtic, Hellenic and maybe Tocharian; see [Krahe 1966](#): 11–12; [Fortson 2010](#): 58–9, 403) is the merger of Proto-Indo-European palatovelar and plain velar plosives into plain velars (PIE **(d)k̑m̑tóm* ‘100’ > PGmc. **hunda-*, Gr. *ἑκατόν*, Lat. *centum*). In contrast, the so-called *satem* branches (Indo-Iranian, Armenian, Balto-Slavic and maybe Albanian; see [Fortson 2010](#): 59) merge Proto-Indo-European labiovelar and plain velar plosives into plain velars and develop the palatovelar plosives further into sibilants (PIE **(d)k̑m̑tóm* ‘100’ > Ved. *śatám*, Av. *satəm*, Lith. *šimtas*). The geographical distribution of *centum* and *satem* branches indicates, however, that only the latter group was truly linguistically innovative. The branches of the peripheral areas thus merely reflect the original situation, with the exception of a trivial merger of palatovelars and plain velars that could easily have happened separately and independently in each branch and, at any rate, must have happened independently in Tocharian vis-à-vis the western *centum* branches.

The *centum*–*satem* distinction aside, scholars have suggested close phylogenetic relationships between Germanic and a range of other languages. The most frequent suggestions set up a Germano-Italo-Celtic unity ([Meillet 1984](#): 131–2; [Porzig 1954](#): 213) or, less frequently, a Germano-Balto-Slavic unity ([Schleicher 1853](#): 787; [Stang 1972](#); also considered as one among several constellations by [Meillet 1984](#): 132 and [Porzig 1954](#): 214). Other scholars venture into larger groupings such as an “alteuropäisch” group consisting of Germanic, Celtic, Italic, Venetic, Illyrian, Baltic and possibly Slavic ([Krahe 1954](#): 48–63; [1962](#): 287–8; [1966](#): 13–14; modified by [Schmid 1968](#)), primarily based on hydronymic evidence; a “North-West Indo-European” group consisting of Italic, Celtic, Germanic and Balto-Slavic ([Oettinger 1997, 1999, 2003](#)); or a general “central” group consisting of Germanic, Balto-Slavic, Indo-Iranian, Armenian, Greek and probably Albanian ([Ringe 2017](#): 6–7). However, these larger groupings are generally based on shared lexical (and derivational) rather than phonological and morphological innovations, which would constitute a more reliable means for establishing linguistic clades (see [Section 2.3](#)). In principle, chances are therefore high that these innovations result from post-split convergence.

In [Sections 10.4.1–7](#), we shall go through the branches with which Germanic is exclusive in sharing specific phonological and morphological features.

10.4.1 *Italic*

Apart from a vast number of lexical innovations, some of which are also shared with Celtic (e.g. Goth. *munþs* ‘mouth’ ~ Lat. *mentum* ‘cheek’, W *mant* ‘jaw’), Germanic shares a handful of innovative phonological and morphological features with Italic ([Porzig 1954](#): 106–17, 123–7; [Krahe 1966](#): 15–17, 20–1).

First among the shared Germano-Italic phonological innovations is the development of PIE **-TT-* > **-ss-* (e.g. pre-PGmc. **uīd-(dʰi)dʰeh₁-t* > PGmc. **wissē(b)* ‘he knew’; PIE **sed-tó-* > Lat. *sessus* ‘seated, sitting’), which may also have been shared with Celtic (Meillet 1984: 57–61; Porzig 1954: 76–8). Second comes the back-vowel quality of the vowel developed in front of Proto-Indo-European syllabic liquids (PIE **l̥, *l̥* > Lat. *or, ol*, Goth. *ur, ul*).

The remaining relevant innovations are morphological. Germanic and Italic show some conformity as regards both the present-stem formation and the function of derived factitive verbs in PIE **-eh₂-je-* (Germanic class II weak verbs ~ Latin 1st conjugation) and stative verbs in PIE **-eh₁-je-* (Germanic class III weak verbs ~ part of the Latin 2nd conjugation, e.g. OHG *dagēn* ‘be silent’ ~ Lat. *tacēre*). Within numeral and adverbial word formation, Germanic and Italic share two innovative derivative suffixes with identical meanings: the creation of distributive numerals from multiplication adverbs by means of the suffix post-PIE **-no-* (**duis-no-* ‘double, of two times’ > ON *tvenmr* ‘double’, Lat. *bīnī* ‘two by two’) and the creation of ablative local adverbs in post-PIE **-tr-ōd* (Goth. *ūtaþro* ‘from outside’; Osc. *contrud* ‘against’).

To the extent that Venetic can be proved to constitute a separate Italic subbranch rather than an independent Indo-European branch (Section 8.2), we note two innovations of Germanic shared with Venetic in this chapter (Porzig 1954: 128; Krahe 1966: 17–18): the addition of post-PIE **g* to the 1sg.acc. of the personal pronoun PIE **mē* ‘me’ due to analogy with the 1sg.nom. **eǵ-* ‘I’ (e.g. Goth. *mik* ‘me’, Ven. *mexo* modelled after Goth. *ik* ‘I’, Ven. *eʒo*) and the creation of an identity pronoun post-PIE **selbʰo-* ‘self’ (Goth. *silba*, Ven. *sselb-*).¹⁷ However, because these two Germano-Venetic innovations are not shared with all Italic subbranches, they must either be independent innovations in Germanic and Venetic or result from convergence between Germanic and Venetic after the initial breakup of Italic.

In a similar vein, granted an Italo-Celtic cladistic node (Section 7.2), the non-participation of Celtic in the Germano-Italic innovations poses serious challenges to the assumption of such a subbranch and suggests that these innovations rather result from secondary convergence after the breakup of Italo-Celtic.

10.4.2 Celtic

Germanic and Celtic had a long period of intensive contact (Porzig 1954: 118–27; Krahe 1966: 18–20; Bousquette & Salmons 2017: 390). Their high number of shared lexical innovations concentrated in certain semantic domains such as

¹⁷ The existence of somewhat similar analogies in the personal pronouns in Anatolian (e.g. Hitt. nom. *uk* ‘I’, acc. *ammuk* ‘me’) and in Greek (*ἐμεγέ*; see Whatmough 2015: 164) strengthens the suspicion that at least this innovation is trivial and may have happened independently in multiple branches (Porzig 1954: 191).

religion and warfare (Hyllested 2009: 117–18, 122) serves as solid evidence thereof. So do a number of indisputable Celtic loanwords in Germanic, e.g. PIE **h₃rēǵ-* ‘king’ > PCelt. **rīg-* ⇒ PGmc. **rīk-*. However, it is often difficult to decide whether a given Germano-Celticism is a shared innovation (or archaism) or reflects a loanword relationship in either direction as exemplified by PIE **h₃reǵ-tu-* > PCelt. **rextu-*, PGmc. **rehtu-* ‘justice’ (Schmidt 1984, 1986; Hyllested 2009: 107).

Notwithstanding the quantity of these lexical isoglosses or their quality for reconstructing a period of Germano-Celtic neighbourhood and convergence, they remain lexical only. Apart from the uncertainties regarding the participation of Celtic in the development of PIE **-TT-* > **-ss-* (Section 10.4.1), Germanic shares no exclusive phonological and morphological innovations with Celtic (Porzig 1954: 123; Hyllested 2009: 108–9). The evidence for a common Germano-Celtic branch is therefore scanty.

10.4.3 Illyrian, Messapic and the Remaining Balkanic Branches

As with both Italic and Celtic, the vast majority of shared innovations between Germanic, on the one hand, and Illyrian and Messapic, on the other, are lexical, e.g. Goth. *þiudans* ‘king’ ~ Illyr. *Teutana* (personal name), but a couple of morphological innovations exists, as well (Porzig 1954: 127–31; Krahe 1966: 18). Only with Illyrian and partially with Greek does Germanic share the generalisation of the *ō*-grade in the declension of feminine *n*-stems (Goth. nom.sg. *tuggo* /tungō/, gen.sg. *tuggons* /tungōns/ ‘tongue’ ~ Illyr. nom.sg. *Aplo*, gen.sg. *Aplōnis* (personal name)). The formation of possessive pronouns with the suffix **-no-* attached to the locative of the personal pronouns is shared with Messapic (e.g. post-PIE **suejino-* ‘his, her’ > Goth. *seins*, Mess. *veinan* (acc.)).

Shared innovations between Germanic and the remaining Balkanic branches of Thracian, Albanian and Hellenic are limited to a handful of lexical correspondences, most of which are also shared with Illyrian (Porzig 1954: 138–9). The only exceptions are the trivial phonological development of PIE **sr* > *str* in Germanic and Thracian-Albanian, which is, however, also shared with Illyrian, Brythonic, Slavic and partly Baltic (e.g. ON *straumr* ‘stream’ ~ Thrac. *Στρώμων* (river name), Illyr. *Straviana*, *Strevintia* (place names), Lith. *strovė* ‘stream’; see Porzig 1954: 78–9; Krahe 1966: 22), and the equally trivial Germanic and Albanian merger of PIE **a* and **o* into **a*, which may also be shared with (Balto-)Slavic (Meillet 1984: 54–6; see also Section 10.4.4).

10.4.4 Balto-Slavic

Most of the innovations shared between Germanic and Balto-Slavic are lexical (e.g. PGmc. **strēla-* ‘arrow’ ~ Lith. *strėlė* ‘arrow, shoot’, OCS *strěla* ‘arrow’;

see esp. Stang 1972 and Nepokupnyj et al. 1989). Four major exceptions from the realm of phonology and morphology come to mind, though (Porzig 1954: 139–47; Krahe 1966: 21–2).

First, and most famously, Germanic and Balto-Slavic agree in forming the dative and instrumental plural with a suffix reflecting a PIE **-m-* rather than the **-b^h-* found in the remaining Indo-European branches (PGmc. dat.pl. **-imiz* as per the Germanic theonyms *Aflims* and *Vatvims* in Roman inscriptions, Lith. dat.pl. *-ms*, instr.pl. *-mis*, OCS dat.pl. *-mъ*, instr.pl. *-mi* ~ Ved. dat.-abl.pl. *-bhyaḥ*, instr.pl. *-bhiḥ*, Lat. dat.-abl.pl. *-bus*, Gaul. dat.pl. *-bo*, Gr. instr.pl. *-φι*; see also Porzig 1954: 90–1). A recent study by Adams (2016: 19–22) indicates that Tocharian belongs to the *m*-group, its ablative ending Toch.B *-mem* reflecting pre-Toch. **-mons*, i.e., the PIE dat.-abl.pl. **-mos* with **n* inserted analogically from the acc.pl. as in OPru. *-mans*. To Olander (2015: 269–70), the **m* of Germanic and Balto-Slavic (and Tocharian) represents a phonological innovation of PIE **-b^hi-* > post-PIE **-m-*. Other scholars, however, regard the *m*-cases as archaic rather than innovative and the **m/b^h* isogloss as a result of different levellings of an original distribution between dative/ablative plural in **m* and instrumental plural in **b^h* (Hirt 1895; Beekes 2011: 188; see also Section 15.4.1).¹⁸

Second, Germanic and Baltic agree on forming the numerals ‘11’ and ‘12’ in a highly non-trivial way by compounding the numerals ‘1’ and ‘2’ with the reflex of PIE **lik^wo-* ‘left’ (Goth. *ainlif* ‘11’, *twalif* ‘12’ ~ Lith. *vieniolika* ‘11’, *dvýlika* ‘12’). The meaning has probably developed along the lines of ‘one left after counting to 10’ (11) and ‘two left after counting to 10’ (12).

The third innovation is phonological. In both Germanic and Balto-Slavic, the inherited vowel qualities PIE **a* and **o* merge into **a*. Since the Slavic development of **a* > *o* is demonstrably late (Meillet 1984: 54), this Germano-Balto-Slavic merger would seem uncontroversial with the short vowels (e.g. PIE **poti-* ‘master’ > Goth. (*bruþ-*)*faþs* ‘bridegroom’, Lith. *pàts* ‘husband, self’; see Meillet 1984: 54–6).¹⁹ However, the Baltic merger of **o* and **a* must postdate Winter’s Law, since PIE **nog^w-* > PBalt. **nōg-* > Lith. *niogas* ‘naked’ (not **nog^w-* > †*nag-* > †*nāg-* > Lith. †*nógas*). The long vowels also require closer investigation. First, the merger of the long vowels only affects parts of the Germano-Balto-Slavic area, since Baltic keeps the reflexes of PIE **ā* and **ō*

¹⁸ For a review of earlier literature on this matter, see Olander (2015: 267–8).

¹⁹ This short-vowel merger also affects Albanian (Section 10.4.3). According to some scholars (e.g. Luraghi 1998: 174), Anatolian partakes, as well, but as Melchert (1993: 251) demonstrates, this merger did not affect Lycian, in which PIE **o* merged with **e* instead of **a*. Thus, it must constitute a secondary shared innovation in Hittite, Palaic and Luwian. In a similar vein, the existence of Brugmann’s Law, which accounts for the different developments of short PIE **a* and **o* in open syllables in Indo-Iranian, witnesses that the identical merger in this branch must also have happened posterior to its separation from the remaining Indo-European branches.

apart (PIE **steh₂* > Lith. *stóti* ‘stand up’ ~ PIE **népōt-* ‘grandson’ > OLith. *nepuotis*). Second, we must accept an intermediary stage of a merged pre-Proto-Germanic **ā* that later develops into PGmc. **ō* as posited in [Section 10.2.1](#). No matter how many branches the mergers of short and long PIE **a* and **o* cover, one fact remains: both mergers represent trivial processes of phonological change and may just as easily have taken place independently in each branch.

Fourth and last, Germanic, Slavic and to some extent Baltic share the equally trivial insertion of **t* into the cluster PIE **sr* with Thracian, Illyrian and Brythonic ([Section 10.4.3](#)).

As a parallel to the case of shared Germano-Italic innovations affecting only the Venetic part of Italic, or only the Italic part of Italo-Celtic ([Section 10.4.1](#)), the fact that Germanic shares innovations with only parts of the Balto-Slavic unity weakens the assumption of an early Germano-Balto-Slavic cladistic node. Being the sole non-trivial innovation shared by all Germanic and Balto-Slavic (and Tocharian?) sub-branches, only the oblique cases in PIE **-m-* may potentially support such an assumption, though with some major potential reservations ([Section 15.4.1](#)). The remaining non-lexical innovations could have either happened independently in each branch or arisen due to convergence at a period when Germanic, Baltic and Slavic had all developed into individual branches. Thus, it is not surprising that Pronk ([Section 15.4.1](#)) dismisses the idea of such a common Germano-Balto-Slavic node.

10.4.5 Armenian

The only innovation uniting Armenian and Germanic is their treatment of the Proto-Indo-European system of plosives. Both branches have undergone ‘consonant shifts’ by changing the articulatory manner of the plosives in similar ways ([Meillet 1984](#): 89–96; [Porzig 1954](#): 80–2; see also [Section 10.2.1](#) for an account of the Germanic developments). The voiced aspirates (PIE **b^h d^h g^h g^{wh}*) developed into unaspirated voiced plosives and/or fricatives), the voiced unaspirated plosives (PIE **b d g g^w*) into unvoiced plosives and, finally, the unvoiced unaspirated plosives (PIE **p t k k^w*) into unvoiced aspirates. These unvoiced aspirates are predominantly retained as such in Armenian (PIE **t k k/k^w* > Arm. *t’ c’ k’*) but have developed further into fricatives in Germanic (PIE **p t k/k k^w* > PGmc. **f p h hw*) and partially in Armenian, too (PIE **p* > Arm. *h*).

Although these developments are indeed substantial, they may still have occurred independently in the two branches in question. As [Meillet \(1984](#): 93–6) mentions, such consonant shifts are trivial innovations with parallels in several other language families worldwide, e.g. Aramaic and some Bantu dialects, and [Porzig \(1954](#): 81–2) questions whether the developments in

Germanic and Armenian are really as parallel as they seem to be at first glance.

10.4.6 Tocharian

The apparent participation of Tocharian in the group of languages that select *m*-variants of the dative/ablative and instrumental plural of case endings (Adams 2016: 19–22; see Section 10.4.4 for a detailed treatment) may position it firmly together with Germanic and Balto-Slavic. Additional parallels between Germanic and Tocharian are limited to lexical elements (Porzig 1954: 97–8, 182–7).

10.4.7 Anatolian

Apart from allegedly both grouping together with Italic and Tocharian in expanding the function of the reflexes of the interrogative pronoun PIE **k^w_{o-}/k^w_{i-}* to include the function of a relative pronoun (Puhvel 1994: 318), Anatolian and Germanic only share lexical isoglosses.²⁰ Even if some among these isoglosses are indeed striking and highly specialised (e.g. ON *herðar* ‘shoulder blades’ ~ Hitt. *kakkartani* ‘shoulder blade’; Goth. *ulbandus* ‘camel’ ~ Hitt. *huwalpant-* ‘hunchback’; Puhvel 1994: 323–4; Melchert 2016: 298–300), they remain lexical and thus less fit for cladistic purposes than phonological and morphological aspects.

10.5 The Position of Germanic

As demonstrated in Section 10.4, no branch offers itself as an obvious candidate for sharing a common node with Germanic in the Indo-European cladistic tree. We could tentatively choose to see the **-m*-variant of the secondary cases (Section 10.4.4) or the collocation of the Germanic 2nd and 3rd classes of weak verbs with the Latin 1st and 2nd conjugation (Section 10.4.1) as evidence in favour of a cladistic partnership with Balto-Slavic and Tocharian or with Italic, respectively. However, these pieces of evidence obviously point in different directions, and as for the Balto-Slavic connection, other pieces of evidence show shared innovations with Baltic only, not with Slavic, which indicates a period of contact and joint development between Germanic and Balto-Slavic

²⁰ The evidence for Germanic partaking in the innovation of expanding the function of the reflexes of the interrogative pronoun is meagre, to say the least. The Germanic languages form their primary relative pronouns in three different ways. East Germanic applies the demonstrative pronoun followed by an enclitic particle *-i*; North Germanic, an indeclinable particle *er* or *es*; and West Germanic, the demonstrative pronoun alone (Krahe 1967: 68–9; see also Porzig 1954: 191).

languages during a relatively late time period and, in any event, after the initial breakup of Balto-Slavic. The same goes for the Germano-Italic innovations that are not also shared with Celtic and thus must postdate the initial breakup of Italo-Celtic. Two linguistic arguments may, however, be presented in favour of a relatively early split of Germanic.

10.5.1 Nominal Ablaut

A well-known, seemingly archaic feature of the Germanic branch is its preservation of Proto-Indo-European nominal ablaut, especially in the heteroclitics. Here we may recall cases such as PGmc. nom. **sōel* (Goth. *sauil*, ON *sól*), obl. **sunn-* (Goth. dat. *sunnin*, ON *sunna*) ‘sun’ < PIE **séh₂-u_{l̥}*, gen. **sh₂-u(é)n-s* and the somewhat parallel PGmc. nom. **fōr* (cf. Goth. *fon*, OHG *fuir*, *fiur* ‘fire’), obl. **fun-* (Goth. gen. *funins*) < PIE **péh₂-u_{r̥}*, **ph₂-u(é)n-s*. With the exception of Anatolian, such nominal ablaut patterns are far less well preserved in the other branches. Although vestiges of these patterns exist throughout the family (Lith. *vanduō* ~ Latv. *udens* ‘water’ < PIE **u(o)d-r/n-* and Lat. *iecur*, gen. *iocineris* ‘liver’ < PIE **iē(ok^w-r/n-*), Germanic appears rather conservative in this respect.

Additional indications for such inherited productivity in Germanic come from a related nominal category, the *n*-stems. There is ample evidence for inherited ablaut patterns in this category, e.g. PIE **krét-ō*, obl. **krit-n-* ‘fever’ (OHG nom. *rido* ~ dat. *riten*); PIE **meh₂k-ō*, obl. **mh₂k-n-* ‘poppy’ (OSw. *val-mōghe* ~ OHG *maho*, *mago*); see further MW *cryd* < PIE **krito-* and Gr. *μῆκον* < PIE **meh₂k-on-*. In other *n*-stems, however, the ablaut appears to be decidedly secondary. A possibly secondary full grade presents itself in, e.g., Nw. dial. *jase* ‘hare’ (< ON **hjasi* < PGmc. **hesan-*) as opposed to pan-Gmc. **hasan-* ~ **hazan-* (OHG *haso*, OE *hara*) and, outside Germanic, Ved. *śása-*, Lat. *cānus* (< **kasno-*) (< PIE **kas-*). Secondary zero grades must in turn be assumed for PGmc. **maþō*, obl. **mutt-* ‘maggot, moth’ (Goth. *maþa* ~ ON *motti*) and **rapō*, obl. **rutt-* ‘rat’ (OHG *rato* ~ MLG *rotte*), apparently from pre-PGmc. **mot-n-* and **(H)rot-n-* (Kroonen 2011: 218–23). The Indo-European nominal ablaut is thus not merely preserved in the Germanic *n*-stems, but seems to have remained productive, a feature long lost in most other branches.

10.5.2 The Preterite-Presents

A second archaic characteristic of Germanic is the retention of the verbal category that is generally held to somehow correspond to the Anatolian *hi*-presents: the Germanic preterite-presents (Section 10.2.2). Examples include

- PGmc. **waita–witume* ‘know’ > Goth. *wait–witum*, ON *veit–vitum*
- PGmc. **maga–magume* ‘can’ > Goth. *mag*, ON *má–megum*
- PGmc. **aiha–aigume* ‘own, have’ > Goth. *aih–aigum*, ON *á–eigum*
- PGmc. **kanna–kunnume* ‘can’ > Goth. *kann–kunnum*, ON *kann–kunnum*
- PGmc. **mana–munume* ‘think’ > Goth. *man*, ON *man–munum*
- PGmc. **skala–skulume* ‘shall, must’ > Goth. *skal*, ON *skal–skulum*

The reconstruction of this category for Proto-Indo-European is debated. Opinions differ as to whether it was a conjugational type of its own or rather originally identical with the perfect (see [Kloekhorst 2018](#) for a discussion).

Regarding the lexical distribution of this class, some of the verbs have parallels in Indo-European languages other than Germanic, e.g. PGmc. **magan-* ~ OCS *mogŏ* (< PIE **mogʰ-* ‘be able’); **munan-* ~ Gk. *μῆμνα* ‘has in mind’ (< PIE **(me-)mon-*); PGmc. **aigan-* ~ Ved. *īṣe* ‘avail over’ (< PIE **(h₂i-)h₂ik-*; see [Hansen 2015](#)); PGmc. **ōgan-* ~ OIr. *ágathar* (< PIE **h₂e-h₂ogʰ-* ‘fear’), yet others are isolated to Germanic, even though they contain more widely attested verbal roots, e.g. PGmc. **kunnan-* (< PIE **ǵneh₃-* ‘know’),²¹ **lisan-* (PIE < **leis-* ‘track’), **ga-nahan-* (< PIE **Hnēk-* ‘reach’) and **skulan-* (< PIE **skel-* ‘owe’). It is tempting to conclude, as a result, that the Germanic preterite-presents, whatever their ultimate origin, were still a productive verbal category when Germanic split off from Proto-Indo-European. This is more reminiscent of the situation in Hittite, where the *hi*-conjugation is still a fully functioning verbal category, than of the situation in the remaining Indo-European branches, where it has largely disappeared and can only be traced through isolated remnants.

10.5.3 Conclusion

Exactly how early Germanic split off remains exceedingly difficult to determine. While Germanic is generally a highly innovative Indo-European sub-branch and lost many of the Proto-Indo-European features still present in Vedic and Greek, the sustained productivity of (1) nominal ablaut and (2) the preterite-presents can be taken as “living fossils”.²² Perhaps then, these are potential indications that Germanic split off from PIE at a relatively early stage, as these features are generally lost in the non-Anatolian branches. Based on this interpretation, we may surmise that Germanic broke off from Proto-Indo-European after Anatolian and just before or after Tocharian.

²¹ The double *n* of **kann-* ~ **kunn-* suggests that it was innovated on the basis of the *neh₂*-present PGmc. **kunnō-* < PIE **ǵnh₃-neh₂-*, which is well-attested outside Germanic (Toch.A *knānat*, Ved. *jānāti*, etc.) and clearly old.

²² The multiply renewed productivity of the root-noun declension type in Germanic ([Hansen 2017](#)) may constitute a third “living fossil” of this type.

References

- Adams, Douglas Q. 2016. Morphosyntax as the handmaid of etymology: On the history of the causal and ablative case-endings in Tocharian. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 15–24. Wiesbaden: Reichert.
- Agee, Joshua R. 2021. Using historical glottometry to subgroup the early Germanic languages. *Journal of Germanic Linguistics* 33. 319–57.
- Andersen, Henning. 2006. Grammatication, regrammatication, and degrammatication: Tense loss in Russian. *Diachronica* 23. 231–58.
- Árnason, Kristján. 2011. *The phonology of Icelandic and Faroese*. Oxford: Oxford University Press.
- Beekes, Robert S. P. 2011. *Comparative Indo-European linguistics: An introduction*. 2nd ed., revised and corrected by M. de Vaan. Amsterdam: Benjamins.
- Bousquette, Joshua & Joseph Salmons. 2017. Germanic. In Mate Kapović (ed.), *The Indo-European languages*. 2nd ed., 387–420. London: Routledge.
- Fortson, Benjamin W., IV. 2010. *Indo-European language and culture: An introduction*. 2nd ed. Chichester: Wiley-Blackwell.
- Hansen, Bjarne Simmelkjær Sandgaard. 2015. The outcome of PIE **#Hi-* and **#Hu-* in Germanic. *Indogermanische Forschungen* 120. 31–76.
- Hansen, Bjarne Simmelkjær Sandgaard. 2016. Lamo talgida or talgida lamo: An indication of a paradigmatic ablative in Proto-Norse? *Arkiv för nordisk filologi* 131. 5–20.
- Hansen, Bjarne Simmelkjær Sandgaard. 2017. Layers of root nouns in Germanic. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 169–82. Wiesbaden: Reichert.
- Henriksen, Carol & Johan van der Auwera. 1994. The Germanic languages. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic languages*, 1–18. London: Routledge.
- Hirt, Hermann. 1895. Über die mit *-m-* und *-bh-* gebildeten Kasussuffixe. *Indogermanische Forschungen* 5. 251–5.
- Hirt, Hermann. 1931. *Handbuch des Urgermanischen*. Vol. 1. Heidelberg: Winter.
- Hoffmann, Karl. 1976. *Aufsätze zur Indoiranistik*. 2 vols., ed. by Johanna Narten. Wiesbaden: Reichert.
- Hyllested, Adam. 2009. On the precursors of Celtic and Germanic. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 21st annual UCLA Indo-European Conference*, 107–28. Bremen: Hempen.
- Jasanoff, Jay H. 1978. Observations on the Germanic *Verschärfung*. *Münchener Studien zur Sprachwissenschaft* 37. 77–90.
- Kloekhorst, Alwin. 2018. The origin of the Hittite *hi*-conjugation. Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 89–106. Ann Arbor: Beech Stave.
- Kluge, Friedrich. 1879. Beiträge zur Geschichte der germanischen Conjugation (Excurs über gotisch *dd* und *gg*). *Quellen und Forschungen* 32. 127–30.
- Kluge, Friedrich. 1884. Die germanische consonantendehnung. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 9. 149–86.
- Kortlandt, Frederik. 1989. The Germanic weak preterite. *Amsterdamer Beiträge zur älteren Germanistik* 28. 101–9.

- Kortlandt, Frederik. 1991. Kluge's Law and the rise of Proto-Germanic geminates. *Amsterdamer Beiträge zur älteren Germanistik* 34. 1–4.
- Kortlandt, Frederik. 2000. The origin of the Goths. *Amsterdamer Beiträge zur älteren Germanistik* 55. 21–5.
- Krahe, Hans. 1954. *Sprache und Vorzeit: Europäische Vorgeschichte nach dem Zeugnis der Sprache*. Heidelberg: Quelle & Meyer.
- Krahe, Hans. 1962. *Die Struktur der alteuropäischen Hydronomie*. Wiesbaden: Akademie der Wissenschaften und der Literatur.
- Krahe, Hans. 1966. *Germanische Sprachwissenschaft*. Vol. 1. *Einleitung und Lautlehre*. 6th ed. Berlin: de Gruyter.
- Krahe, Hans. 1967. *Germanische Sprachwissenschaft*. Vol. 2. *Formenlehre*. 6th ed. Berlin: de Gruyter.
- Kroonen, Guus. 2011. *The Proto-Germanic n-stems: A study in diachronic morphophonology*. Amsterdam: Rodopi.
- Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden: Brill.
- Kummel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.
- Lehmann, Winfred P. 1966. The grouping of the Germanic languages. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 13–27. Berkeley: University of California Press.
- Luraghi, Silvia. 1998. The Anatolian languages. In Anna Giacalone Ramat & Paolo Ramat (eds.), *The Indo-European languages*, 169–96. London: Routledge.
- Lühr, Rosemarie. 1984. Reste der athematischen Konjugation in den germanischen Sprachen: zu 'sein' und 'tun'. In Jürgen Untermann & Bela Brogyanyi (eds.), *Das Germanische und die Rekonstruktion der indogermanischen Grundsprache*, 25–90. Amsterdam: Benjamins.
- Lühr, Rosemarie. 1988. *Expressivität und Lautgesetz im Germanischen*. Heidelberg: Winter.
- Markey, Thomas L. 2001. A tale of two helmets: The Negau A and B inscriptions. *Journal of Indo-European Studies* 29. 69–172.
- Maurer, Friedrich. 1942. *Nordgermanen und Alemannen: Studien zur germanischen und frühdeutschen Sprachgeschichte, Stammes- und Volkskunde*. Strasbourg: Hünenburg.
- Maurer, Friedrich. 1952. *Nordgermanen und Alemannen: Studien zur germanischen und frühdeutschen Sprachgeschichte, Stammes- und Volkskunde*. 3rd ed. Bern: Francke & Lehnen.
- Meid, Wolfgang. 1971. *Das germanische Präteritum: Indogermanische Grundlagen und Ausbreitung im Germanischen*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Meillet, Antoine. 1984 [1908]. *Les dialectes indo-européens*. New expanded edition with introduction. Paris: Honoré Champion.
- Melchert, H. Craig. 1993. Historical phonology of Anatolian. *Journal of Indo-European Studies* 21. 237–57.
- Melchert, H. Craig. 2016. "Western affinities" of Anatolian. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Etymology and the European lexicon*, 297–305. Wiesbaden: Reichert.
- Must, Gustav. 1957. The problem of the inscription on Helmet B of Negau. *Harvard Studies in Classical Philology* 62. 51–9.

- Nepokupnyj, A. P., et al. 1989. *Obščaja leksika germanskix i baltoslavjanskix jazykov* [The shared lexicon of Germanic and Balto-Slavic]. Kiev: Naukova dumka.
- Nielsen, Hans Frede. 1981. Krimgotisk: oversigt og problemstillinger. In Hans Frede Nielsen (ed.), *Sammenlignende studier i gotisk*, 283–99. Odense: Odense Universitet.
- Nielsen, Hans Frede. 2000. *The early runic language of Scandinavia: Studies in Germanic dialect geography*. Heidelberg: Winter.
- Noreen, Adolf. 1894. *Abriss der urgermanischen Lautlehre: Mit besonderer Rücksicht auf die nordischen Sprachen*. Strasbourg: Trübner.
- Nørgård-Sørensen, Jens & Lars Heltoft. 2015. Grammaticalisation as paradigmatisation. In Andrew D. M. Smith, Graeme Trousdale & Richard WALTERIT (eds.), *New directions in grammaticalisation research*, 261–92. Amsterdam: Benjamins.
- Oettinger, Norbert. 1997. Grundsätzliche Überlegungen zum Nordwest-Indogermanischen. *Incontri Linguistici* 20, 93–111.
- Oettinger, Norbert. 1999. Zum nordwest-indogermanischen Lexikon (mit einer Bemerkung zum hethitischen Genitiv auf -l). In Peter Anreiter (ed.), *Studia Celtica et Indogermanica: Festschrift für Wolfgang Meid zum 70. Geburtstag*, 261–7. Budapest: Archaeolingua Alapítvány.
- Oettinger, Norbert. 2003. Neuerungen in Lexikon und Wortbildung des Nordwest-Indogermanischen. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in prehistoric Europe*, 183–93. Heidelberg: Winter.
- Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. Leiden: Brill.
- Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. Heidelberg: Winter.
- Puhvel, Jaan. 1994. West-Indo-European affinities of Anatolian. In George E. Dunkel et al. (eds.), *Früh-, Mittel-, Spätindogermanisch*, 315–24. Wiesbaden: Reichert.
- Rasmussen, Jens Elmegård. 1990. Germanic *Verschärfung*: Tying up loose ends. In Henning Andersen & Konrad Koerner (eds.), *Historical linguistics 1987*, 425–41. Amsterdam: Benjamins. Reprinted in Rasmussen 1999, 375–93.
- Rasmussen, Jens Elmegård. 1996. On the origin of the Germanic weak preterite. *Copenhagen Working Papers in Linguistics* 4, 161–8. Reprinted in Rasmussen 1999, 597–603.
- Rasmussen, Jens Elmegård. 1999. *Selected papers on Indo-European linguistics: With a section on comparative Eskimo linguistics*. Copenhagen: Museum Tusculanum.
- Ringe, Don. 2017. *From Proto-Indo-European to Proto-Germanic*. 2nd ed. Oxford: Oxford University Press.
- Robinson, Orrin W. 1992. *Old English and its closest relatives: A survey of the earliest Germanic languages*. London: Routledge.
- Rösel, Ludwig. 1962. *Die Gliederung der germanischen Sprachen nach dem Zeugnis ihrer Flexionsformen*. Nuremberg: Carl.
- Schleicher, August. 1853. Die ersten Spaltungen des indogermanischen Urvolkes. *Allgemeine Zeitung für Wissenschaft und Literatur* 3, 786–7.
- Schleicher, August. 1860. *Die deutsche Sprache*. Stuttgart: Cotta.
- Schmid, Wolfgang P. 1968. *Alteuropäisch und Indogermanisch*. Wiesbaden: Akademie der Wissenschaften und der Literatur.

- Schmidt, Karl Horst. 1984. Keltisch und Germanisch. In Jürgen Untermann & Bela Brogyanyi (eds.), *Das Germanische und die Rekonstruktion der indogermanischen Grundsprache*, 113–54. Amsterdam: Benjamins.
- Schmidt, Karl Horst. 1986. Keltisch-germanische Isoglossen und ihre sprachgeschichtlichen Implikationen. In Heinrich Beck (ed.), *Germanenprobleme in heutiger Sicht*, 231–47. Berlin: de Gruyter.
- Schwarz, Ernst. 1951. *Goten, Nordgermanen, Angelsachsen: Studien zur Ausgliederung der germanischen Sprachen*. Bern: Francke.
- Stang, Christian S. 1972. *Lexikalische Sonderübereinstimmungen zwischen dem Slavischen, Baltischen und Germanischen*. Oslo: Universitetsforlaget.
- Stifter, David. 2009. The Proto-Germanic shift $*\bar{a} > *\bar{o}$ and early Germanic linguistic contacts. *Historische Sprachforschung* 122. 268–83.
- Streitberg, Wilhelm. 1896. *Urgermanische Grammatik*. Heidelberg: Winter.
- Whatmough, Joshua. 2015. *The foundations of Roman Italy*. London: Routledge.

Lucien van Beek

11.1 Introduction

Greek is one of the earliest attested languages of the IE family, starting with Mycenaean in the fourteenth–twelfth century BCE (on the dating of the tablets, see [Driessen 2008](#)). From the so-called Dark Ages (twelfth–ninth century BCE), we have only one written piece of evidence in Greek (Cypriot *O-pe-le-ta-u*, perhaps mid-tenth century). Starting in the eighth century BCE, alphabetic inscriptions appear in various different dialects and from all corners of the Greek world; moreover, literary Greek starts with the Homeric epics.

From the Mycenaean period onwards, Greek was spoken in the southernmost parts of the Balkan peninsula (Epirus, Thessaly, and further south) and on the islands in the Aegean (Crete, Cyclades) and Ionian seas. Processes of migration and colonization starting as early as the Mycenaean period brought Greek across the Aegean to the Western and Southern Asia Minor coastline, to Cyprus and probably the Levant, and from the eighth century onwards to Sicily, the Italic peninsula, the Rhone delta, Libya, Egypt, and the Black Sea region.

Mycenaean Greek was written in a syllabic script (Linear B). With the destruction of the palaces, Linear B went out of use, but on Cyprus a related syllabary survived, most inscriptions dating to the eighth–fourth century BCE. All other first-millennium varieties of Greek were written in different local forms of the Greek alphabet, which was adopted from the Phoenician *abjad* during the Dark Ages (the exact date(s) and place(s) of adoption are still debated).¹

Ancient Greek is attested in many (at least thirty) different dialects: from the beginning of the Dark Ages until the Classical period, almost every *polis* had its own local (epichoric) variety and local alphabet, reflecting the political fragmentation of Greece. Broadly speaking, the following dialects are attested in the inscriptional record (cf. [Buck 1955](#)), divided into four main groups (see [Section 11.3](#)):

This chapter was made possible by a VENI grant from NWO (Netherlands Organization for Scientific Research) for the project *Unraveling Homer's language*.

¹ An eleventh-century date has recently been proposed ([Waal 2018](#)).

- Arcado-Cypriot: Arcadian (Central Peloponnese) and Cypriot (Cyprus); Mycenaean is closely related to both dialects
- Ionic-Attic: Attic (Attica), Western Ionic (Euboea, Oropos), Central Ionic (Cycladic islands), Eastern Ionic (Chios and the Asia Minor coastline from Smyrna to Halicarnassus)
- Aeolic: Thessalian (Thessaly, with five regional varieties), Boeotian (Boeotia), and Lesbian/Aeolic proper (Lesbos and the Western Asia Minor coastline north of Smyrna)
- West Greek, usually subdivided into Doric and North-West Greek dialects (cf. [Mendez Dosuna 2007b](#)):
 - Doric dialects were spoken around the Saronic Gulf (Megarian, Corinthian, Eastern Argolic), on the Peloponnese (Western Argolic, Laconian, Messenian), on the southern Aegean islands (Cretan, Theran, Dodecanese (Cos, Rhodes)), and the Ionian islands (including Corcyrean).
 - North-West Greek dialects were spoken North of the Gulf of Corinth: Locrian, Phocian, Delphic, Acarnanian, Aetolian, Epirotic.²
 - The dialect of Elis has many peculiar features; that of Achaea is marginally attested.
 - Various West Greek dialects were transported to colonies in Magna Graecia, where they developed local characteristics (Syracusan from Corinthian, Tarentine and Heracleon from Laconian, etc.); Cyrenaean developed from Theran.

Pamphylian (around present-day Antalya, southern coast of Asia Minor) is fragmentarily attested and difficult to classify ([Brixhe 1976; 2013](#)).

A linguistic description of most dialects, however, is hampered in various ways (for a detailed methodological discussion see [García Ramón 2017](#)). First, there are large chronological and geographic gaps in the often fragmentary attestations of most dialects. In the archaic period, longer inscriptions (e.g. the Gortyn Law Code) are scarce, and there are not any longer dialect texts from Messenia, Achaea, and large parts the North-Western realm. Secondly, the range of subjects covered in prose inscriptions is narrow (mostly treaties and regulations), and the language is often formulaic or standardized. This may also hold for Mycenaean, where the relative lack of variation between different find spots is suggestive of a bureaucratic register. Third, a tendency toward koineization starts relatively early in most areas, and the tendency to actively promote local dialect peculiarities in official inscriptions led to hyper-dialectal forms. Finally, even with the dialects that are known well (Classical Attic and, to some extent, Eastern Ionic), it must be taken into account that literary texts do not always reflect the actual linguistic situation.

² Many of these dialects are only fragmentarily attested.

Indeed, utilizing forms of literary Greek poses problems of a different nature. Most archaic forms of poetry are not in local dialect, but in genre-dependent (epic, lyric, drama, etc.) linguistic forms. Specific features became established as markers of certain genres (e.g. feminine participles in *-οισα* in choral lyric, probably reflecting the prestige of Lesbian poetry). Moreover, all genres share a considerable body of archaic grammatical and lexical features that were absent from most vernaculars. These features may derive from a traditional poetic language (a “poetic Koine”) with roots in the late second millennium.

For these reasons, it is often difficult to assign features attested in literary texts to a specific dialect. Thus, alongside contemporary Lesbian forms, the language of Sappho and Alcaeus contains common poetic forms, borrowings from Ionic and from epic, and probably also artificial forms.³ Epic Greek has a general Ionic phonological veneer and contains many specifically Ionic grammatical and lexical features. However, as the traditional language of verse-composition in hexameters, it also contains large numbers of archaic words, morphemes, and phrases. Some of these can be assigned to dialects other than Ionic (Aeolic, probably also Mycenaean), but often dialect assignment is difficult. Finally, a considerable number of typical Homeric forms are artificial creations (for an overview, see [Hackstein 2010](#)).

11.2 Evidence for the Greek Branch

This section aims to present all innovative developments (including significant choices between alternatives) that set Proto-Greek apart from other branches.⁴ In combination with the virtual absence of demonstrably old divergences between the Greek dialects, this enumeration shows that Proto-Greek existed as a real prehistoric linguistic entity, thus disproving Garrett’s provocative claim that there are hardly any “demonstrable and uniquely Proto-Greek innovations in phonology and inflectional morphology” (2006: 141).⁵

First, some remarks concerning relative chronology. The Mycenaean evidence allows us to assign certain changes to the period after the adoption of Linear B (e.g. **p₂* > *pt*, or the lenition of initial *yod*). It is not always easy, however, to distinguish between Proto-Greek innovations and later shared Common Greek developments. An often-cited example is the introduction of **-wot-* as the perfect participle suffix. This innovation was formerly reconstructed for Proto-Greek because it occurs in all first-millennium dialects (except for Aeolic, which uses the suffix **-ont-*), but Mycenaean shows that Proto-Greek retained **-woh-*. However, although the Proto-Greek status of some of the individual changes below may be doubted, it

³ An extensive treatment is [Bowie 1981](#).

⁴ For a similar but less extensive list, see [Clackson 2007](#).

⁵ Exaggerated doubts concerning our ability to reconstruct Proto-Greek also surface in Risch’s work (e.g. [Risch 1963](#)).

is clear that they all took place between PIE and attested Greek; hence, the majority will have taken place before the split into North and South Greek.

11.2.1 Phonological Innovations Shared by All Greek Dialects

- Specific laryngeal vocalizations, including
 - word-initial before consonant plus vowel (**HCV*-): triple reflex *e, a, o*⁶
 - word-initial before resonant plus consonant (**HRC*-): triple reflex *e, a, o*
 - between two consonants (**CHC*): triple reflex *e, a, o*; this probably included word-initial **RHC*-, cf. *μακρός* ‘long’ < **mh₂k-ró-* beside *μήκιστος, μήκος*
 - *CRHC* > PGr. /CRēC/, /CRāC/, /CRōC/⁷
 - *CRHV* > PGr. /CaRV/ (with coloring of V by the laryngeal)⁸
 - the development of **CiHC* and **CuHC* remains disputed: *θῦμός* ‘spirit’ < **d^huh₂mó-* ‘smoke’ (Lat. *fūmus*, Ved. *dhūmá-*, also Hitt. *tuhhuwai-*, all ‘smoke’) is a certain example of a long-vocalic reflex. On the other hand, Ved. *jīvati, jīvā-* and Lat. *vīvō, vīvus* seem to imply a vocalization **CīōC* < **Cīh₃C* for the cognate formations *ζῶω* ‘live’, *ζωός* ‘alive’
 - *-ih₂* > *-ia* at word end (nom.sg. of the fem. motion suffix), also **-ih₁* > *-ie* (only in dual **h₃ek^w-ih₁* > Hom. *ὄσσε* ‘eyes’); it is debated whether this change was phonetically regular or analogical.
- The double reflex of **i-*, which merges with **di-* (plus **gi-*, **g^wi-*) in one subset of lexemes that have correspondences with **i-* in other IE languages (e.g. *ζέω* ‘boil’, Myc. *ze-so-me-no*; *ζυγόν* ‘yoke’ and *ζεύνυμι* ‘connect’, Myc. *ze-u-ke-si*), but was retained and developed into *h-* in another subset (relative pron. *ὅς*, Myc. *jo-*, *o-* beside Ved. *yáḥ*; *ἥπαρ* ‘liver’ beside Lat. *iecur*). The distribution between both reflexes, which is the same in all Greek dialects (including Mycenaean), represents an exclusive common innovation of Proto-Greek. The exact conditioning factor, probably the

⁶ The divergent initial reflex of Doric *φικατι* ‘twenty’ ~ Classical *εἴκοσι* < **h₁mi-Hk^hmt-i* (with problematic *o* < **m*) is unexplained, but this does not suffice to show that the laryngeals were retained until after PGr.

⁷ The divergent form *πρῶτος* vs. West Greek *πρῶτος* of the ordinal ‘first’ must reflect a contracted superlative PGr. **pro-ato-* (cf. Cowgill 1970: 123, 148). There is some evidence for a disyllabic reflex of **CRHC*: *πρᾶχός* ‘rough’ < **d^hrh₂g^h-u-*, *θράσσω* ‘stir’ < **d^hr(e)h₂g^h-*, but *ταράσσω* ‘id.’ < **d^hrh₂g^h-*. It is often claimed that the disyllabic treatment occurred only when the liquid was accented (e.g. Rix 1992: 73), but in my view this is uncertain. Another plausible possibility is that the disyllabic reflex was regular before /CC/, while the long vowel reflex occurred before /CV/ (van Beek 2021a).

⁸ **CRh₃V-* may have yielded PGr. /CoRV/, with rounding of the anaptyctic vowel caused by the following labio-laryngeal (cf. *μολεῖν* ‘come’, *πορεῖν* ‘give’ < **mlh₃-e/o-*, **prh₃-e/o-*). The Lesbian form *χόλαισι* ‘are slack’ (Alcaeus) corresponding to Classical *χαλῶσι* (*χαλάω*) is not sufficient evidence for positing a distinct reflex for Aeolic (pace Peters 1980: 28).

presence or absence of an initial laryngeal (cf. [García Ramón 1999](#)), is still disputed (cf. [van Beek 2019](#)).

3. Loss of word-final stops, including stop clusters (voc. *ἄνα* ‘lord’ < **wanakt*).
4. Restrictions on allowed stop clusters, including developments of “thorn clusters” (two consecutive stops are allowed only if the second stop is dental, e.g. *κτ* or *πτ*; while *τπ**, *τκ**, *κπ**, *πκ** are disallowed). This situation is pre-Mycenaean in view of e.g. *e-qi-ti-wo-e* /ek^{wh}t^hiwohes/ perf.ptc. ‘perished’ from PIE **d^hg^{wh}ei-*.
5. Development of voiceless aspirates /t^h k^h p^h k^{wh}/ from the PIE “mediae aspiratae”, already completed in Mycenaean (cf. *te-o* /t^he(h)os/ ‘god’ from PIE **d^hh₁s-ó-*; but contrast [Section 11.4](#) on Macedonian and Phrygian).
6. Development of a circumflex accent: the pitch on long vowels may fall on the first mora (circumflex accent) or on the second mora (acute accent). The distinction was probably phonologized when early contractions took place, not long after the loss of intervocalic laryngeals (e.g. *τιμή* gen.sg. < **-éh₂-os* vs. *τιμή* nom.sg. < **-éh₂*).
7. The Law of Limitation: the pitch accent can be assigned only to the last four morae of a prosodic word, and only to the last three morae if the final syllable is accentually long.
8. Lenition **s* > *h* in different positions: (a) word-initially before vowels or *R* (= any liquid, nasal, or glide); (b) between vowels and in the intervocalic clusters **-sR-* and **-Rs-* (probable exception: *-rs-* and *-ls-* were not lenited if the directly preceding syllabic nucleus carried the accent).
9. The syllabic nasals yielded a nasal vowel [ã] or [ə̃] in Proto-Greek. This normally merged with /a/ in all dialects, but in some dialects we also find /o/ under specific, yet still uncertain, conditions (perhaps in a labial environment).⁹
10. Cowgill’s Law, i.e. **o* > *u* in certain environments involving labials and nasals. In various words this raising occurs in all Greek dialects, e.g. *νύξ* ‘night’ < **nok^wt-*. However, not all dialects show this raising in the same words (cf. Ion.-Att. *ὄνομα* vs. Dor. Aeol. *ὄνομα*), and the conditions are still in part uncertain; see [Vine 1999](#).

The laryngeal changes under (1) are mostly specific to Greek, but some are shared with Phrygian ([Section 11.4.2](#)). This may also hold for developments (3) and (4), which are equally attested in Phrygian, although the Greek loss of final stops is difficult to date (the Linear B syllabary does not make it possible to determine whether they were present in Mycenaean or not; contrast also Phryg. voc. *-vanak* with Gr. *ἄνα* ‘Lord!’). The Law of Limitation is difficult to date as we have no evidence for accentuation in most dialects.

⁹ Discussion of the evidence in [Thompson 1996–7](#): 316–20.

A development **CRHV-* > **CaRV-* is also found in Italic and Celtic, but it is probably independent, as in those branches *a*-coloring of anaptyctic schwa is unsurprising. The vocalization in (9) may be independent of that in Indo-Iranian, as the Greek outcomes /a/, /o/ postdate the Graeco-Phrygian stage (**ŋ* > Phr. *an*).

Certain developments involving clusters of stop plus glide are also likely to be Proto-Greek:

11. Intervocalic **t^(h)i* merges with PGr. **ts* (Ion.-Att. μέσος < **med^hios*, τόσος < **totios*; Arc. μέσος; Myc. *to-so*; Boeot. μέττος; most other dialects μέσσος; older Cretan may preserve /ts/). In productive formations, **t^(h)i* was restored; its reflex merged with that of **k^(h)i* in most dialects but not in Mycenaean.

11.2.2 Morphological Innovations: Verbal Stem Formation and Endings

12. Development of an aorist in *-θη-*, in addition to the inagentive aorist in *-η-* (which reflects “stative” **-eh₁-*). The exact origin and genesis of this formation are still disputed.
13. Creation of a *κ*-perfect, where *-κ-* was originally found only in the indic. sg.¹⁰ Greek productively extended this morpheme (perhaps originally an aorist marker, cf. unreduplicated Lat. *fēcī*, *iēcī* beside *έθηκα*, *έηκα*), first to intransitive perfects of long-vocalic roots (e.g. *πέφδκα*, *έστηκα*), later also to transitive perfects (e.g. *λέλδκα*) and other stem types.
14. Replacement of the perf.act.3pl. ending **-ēr* with **-nti*, reflected as *-ατι* in WGr. dialects and as *-ασι* in Arcadian (Buck 1955: 112). This ending was later adapted to **-anti* (> Att.-Ion. *-ασι*) in most dialects.
15. The “alpha-thematic” sigmatic aorist paradigm, which was based on the 1sg. after the word-final change **-m* > *-a*; the 3sg. received the thematic ending *-e* after the loss of **-t*.
16. Replacement of the stative endings by the middle endings 3sg. *-to*, 3pl. *-nto*.
17. Creation of new secondary middle endings 1sg. **-mān* (unique to Greek) and 2sg. **-so* (as in other branches, including Italic and Germanic).¹¹
18. Creation of primary middle endings in *-i*.
19. Development of a medio-passive perfect stem (see Section 11.4.2).
20. Creation of an active pluperfect with a suffix **-e-* and alpha-thematic endings (Hom. *έπεποιθεα*).¹²

¹⁰ Cf. Att. *έστώς*; *τεθνεώς* beside *έστηκα*; *τέθνηκα*.

¹¹ It cannot be excluded, however, that the PIE stative endings 1sg. **-h₂*, 2sg. **-th₂o* were originally distinct from middle **-mh₂*, 2sg. **-so*. Cf. Kortlandt 1981.

¹² However, the antiquity and spread of this formation are difficult to assess. The irregular Homeric pluperfect *ήδη* ‘knew’ is certainly old; it has been compared with PCelt. **wēdī* < **ueid-eh₁-* by Schrijver (1999).

21. Certain productive reduplication patterns:
 - a. default *Ce-* (perfect stem), *Ci-* (present stem) for roots with simple onsets
 - b. “Attic reduplication” in roots starting with a vowel (e.g. *ἐλνθ-* → *ἐλνθλνθ-*)
 - c. full reduplication in roots of the structure /VC-/ (e.g. *ἀρ-* → *ἀρηρη-*)
 - d. /e-/ in the perfect of roots with complex onsets (e.g. perf.mid. *ἔξευγμαi*).
22. The infinitive endings:
 - a. thematic **-e-hen* (e.g. Myc. *e-ke-e* /ek^hehen/ ‘have’, Att. *-ειν*, etc.)
 - b. athematic **-men*, **-menai* (Lesb. *ἔμμεναι* ‘be’) and **-hen* (Myc. *te-re-ja-e* /teleiāhen/ ‘fulfill’), **-henai* (Att. *ἔνναι* ‘go’)¹³
 - c. **(t)sai* (s-aorist)
 - d. **-st^hai* (middle).
23. Creation of a denominative factitive class in PGr. *-ō-* (type *δηλώω*), see [Tucker \(1990\)](#).

11.2.3 Morphological Innovations: The Cases, Nominal Endings, and Nominal Stem Formation

24. The PGr. dat.-loc.pl. ending *-si* (for PIE **-su*) arose by analogical introduction of *-i* from the loc.sg. ending, probably aided by instr.pl. **-b^hi*.¹⁴
25. Case syncretism: Proto-Greek merged the dative and locative plural of all declensions (PGr. **-oisi*, *-āsi*, *-sī*).
26. Greek has various clitics and suffixes marking spatial relations: **-de* cliticized to the accusative of direction, e.g. *οἶκόνδε* ‘home’ (already Mycenaean), **-t^hi* (locative, e.g. *οἶκοθι* ‘at home’), **-t^hen* (ablative, e.g. *παντόθεν* ‘from all sides’), but also local **-t^hη > -θα* as well as **-t^he* after local adverbs; at least **-t^hi* and **-t^hen* originated in adverbial pronouns (cf. *πόθι* ‘where’, *πόθεν* ‘whence’) and were innovations of Proto-Greek.

Proto-Greek had more innovations (e.g. the introduction of nom.pl. endings *-oi*, *-ai* in the first and second declension, the extension of the 3rd decl. n.pl. ending *-ǎ < *-h₂* to thematic stems replacing the reflex of **-eh₂*, or the generalization of the 3rd decl. gen.sg. ending *-os* to the exclusion of **-es*). However, since most of them are shared with various different other branches and fairly trivial developments, they cannot be utilized for purposes of subgrouping.

In nominal stem formation, innovations include:

27. The suffixes *-ēu-* (masculine persons or professions), *-ād-* and *-id-* (denoting appurtenance).

¹³ For the relation between **-hen(ai)* and **-men(ai)*, see [van Beek in press](#). The suffixes **-men-* and **-hen-* could both be extended with *-ai* under certain specific conditions. In Lesbian, *-μεναι* occurs only with monosyllabic stems containing a short vowel.

¹⁴ [Pace Garrett \(2006: 140\)](#), this is not “a trivial adaption”.

28. The extended form in *-t-* (Classical *-ματ-*, *-ατ-*) of the suffixes **-mḡ-*, **-r/-n-* in neuter nouns.
29. The extended form of the comparative suffix **-is-on-* > *-ίων* (unattested in Myc., though).
30. The use of **-tero-* as a comparative suffix with gradable adjectives.
31. The superlative suffix **(t)mto-* > *-(τ)ατος*, replacing **(t)mHo-* (cf. Lat. *intimus* ‘innermost’, Ved. *ántama-* ‘nearest’).

11.2.4 Pronouns

32. Acc.pl. of the personal pronouns in *-mé* (generalized orthotonic forms **ḡs-mé*, **us-mé*).
33. Reshaping of the nom.pl. **mei*, **ius* of the personal pronouns after the acc.: **ḡsm-es*, **usm-es* (cf. Dor. *ἄμέες*, *ὕμέες*; Aeol. *ἄμμεες*, *ὕμμεες*).
34. The dative of personal pronouns in *-i(n)*: clitic Ion.-Att. *ἡμιν*, orthotonic Dor. *ἡμίν*, Lesb. *ἄμμι(ν)* (contrast Ved. dat. *asmé* < **-me-i*).
35. Creation of a stem form *σφε-* beside *σφι(ν)* ‘to them(selves)’, probably a clitic form of PIE **se-b^hei*.
36. Grammaticalization of anaphoric/demonstrative *οὗτος*, *αὕτη*, *τοῦτο* (intermediate deixis) from **só* (*h₂*)*u* plus **to-* (the first part corresponds to Ved. *sá u* and the nom.sg. pronoun Plr. **hau* (OAv. *huuō* /*hau*/, OPers. *hauv*), Ved. *asáu*).
37. Creation of the demonstrative *κεῖνος* / *ἐκεῖνος* (distal deixis).
38. Reflexive *αὐτός* ‘same; self’, grammaticalized from **h₂eu* ‘again’ plus anaphoric-demonstrative **to-*.
39. Creation of a negation *οὐκ(i)*, *οὐ*, probably from **(ne)* ... **h₂oju* *k^{wid}* (Cowgill 1960).

11.2.5 The Lexicon and Remaining Innovations

Lexical innovations are more difficult to utilize for the purpose of subgrouping, but they may complement the picture gained from the phonological and morphological innovations. Some typical lexical innovations of Greek are (a full list would be much longer):

40. The verb ‘wish, choose’ has a root PGr. **g^wel-* or **g^wol-* instead of PIE **uelh₁-* (βούλομαι, Arc., Eub. βόλομαι, Thess. βέλλομαι, WGr. δειλομαι, etc.).
41. The verb ‘die’ has the root PGr. **t^hnā-*, **t^hana-*.
42. The word for ‘guest, stranger’ is PGr. **ksenwo-*.

A large amount of the Greek lexicon was borrowed from the indigenous language(s) of the Hellenic peninsula. Beekes (2014) views this as one single non-Indo-European language which he calls “Pre-Greek”, but while the Greek lexicon indeed has an important non-Indo-European element, it is difficult to

determine when, where, and from how many different varieties this material was taken. The forms *πίργος* ‘fortification’ < *b^h(o)rǵ^h- and *τύμβος* ‘grave’ < *d^h(o)mb^h- presuppose an Indo-European donor language.

11.3 The Internal Structure of Greek

The Ancients distinguished four main dialects of Greek: Attic, Ionic, Doric, and Aeolic. As they recognized that Attic and Ionic were very closely related, a basic three-way distinction is implied (also reflected in the three Hellenic tribes and their ancestors *Δῶρος*, *Ξοῦθος*, and *Αἰόλος* in Hesiod fr. 9 M-W). However, ancient scholarship was interested mainly in literary languages, not in spoken dialects (see [Tribulato 2019](#)).

After the decipherment of the Cypriot syllabary, however, scholars quickly realized that Arcadian and Cypriot were much more closely related to each other than to Thessalian and Boeotian, and that the Ancients used “Aeolic” as a catchall term for anything that was not Ionic, Attic, or Doric. Even so, the threefold distinction (and the inclusion of Arcado-Cypriot among the Aeolic dialects) was largely maintained.¹⁵ In fact, the theory that Ionians, Aeolians, and Dorians existed as distinct ethnic and linguistic groups as early as 2000 BCE, and that they migrated into the Hellenic peninsula in three chronologically distinct waves (Kretschmer’s *Wellentheorie*), held sway for a long time.

This picture was changed radically by two landmark studies, [Porzig 1954](#) and [Risch 1955](#); see also [Risch 1963](#). Both scholars independently showed that Arcado-Cypriot was a distinct dialect group with close genetic ties to Ionic-Attic. Moreover, both argued that Asia Minor Aeolic (Lesbian) had been influenced substantially by neighboring Ionic dialects, and that East Thessalian is the most conservative Aeolic dialect. In addition, Risch made a plausible argument for reconstructing a first split into North Greek and South Greek (comprising Arcado-Cypriot and Ionic-Attic) in the early second millennium.¹⁶ It is now widely accepted that South Greek is characterized by the following exclusive innovations:

- assibilation *t^hi > /si/ (e.g. 3sg. *δίδωσι*)
- simplification PGr. *ts and *ss > s, also after short vowels (e.g. *μέσος*)¹⁷

¹⁵ For a good summary of earlier works on Greek dialect classification and subgrouping, see [Morpurgo Davies 1992](#).

¹⁶ Many scholars still use the terms West Greek and East Greek (cf. [Porzig 1954](#)) instead of Risch’s North Greek and South Greek, respectively. In order to avoid confusion, I stick to Risch’s terminology and reserve “West Greek” for the dialect group that comprises all Doric and Northwest Greek dialects.

¹⁷ According to Risch, *ts > s fed the assibilation *ti > si, but the antiquity of (*ts >) ss > s cannot be proven because Linear B does not write geminates (Myc. *to-so* corresponding to Ion.-Att. *τόσος*).

- athematic infinitives **(h)én*, **(h)énai* (Dor. and Aeol. *-μεν*, *-μεναι*)¹⁸
- correlative temporal adverbs in */-te/*, e.g. *τότε* ‘then’ (Aeol. *-τα*, Dor. *-κα*)
- temporal conjunction *εἰ* (Dor. Aeol. *αἰ*), but Cypr. has *e-*
- nom.pl. *τοί*, *ταί* of the demonstrative replaced by *οἱ*, *αἱ* (probably also Aeolic).

There are few (if any) old innovations that are characteristic for all North Greek dialects. The best candidate is the *e*-vocalism of the present stem ‘want’ (Thess. *βέλλομαι*, WGr. *δείλομαι*, etc.), but it remains uncertain whether this is a shared innovation rather than an archaism. It is likely that certain distinctive Aeolic innovations occurred between the separation of South Greek and the twelfth century (Section 11.3.7).

Following Risch, we may distinguish three periods:

- a. Mycenaean period (relative stability, probably increasing local differentiation)
- b. Dark Ages (high mobility; rapid language change, convergence)
- c. ninth century BCE until the Classical period (the dialects occupy their historical locations; colonization movements; increasing local differentiation).

Various linguistic innovations can be assigned to one of these periods, based on (1) relative chronology, (2) linguistic geography, and (3) their presence or absence in Mycenaean.¹⁹

11.3.1 *Mycenaean*

Mycenaean is clearly a South Greek dialect, as evidenced by the assibilation of voiceless dental stops (e.g. *di-do-si* /didonsi/ ‘they give’), the conjunction *o-te* ‘when’, and an athematic infinitive in */-hen/* (*te-re-ja-e* /teleiāhen/ ‘fulfill’).

Apart from this, however, the position of Mycenaean relative to the first-millennium dialects is less clear.²⁰ Arcadian and Cypriot are closely related dialects, but it must be borne in mind that most exclusive Arcado-Cypriot innovations are not attested in Linear B (see below). An exception in this respect might be Myc. *pe-i* /sp^hehi/, an innovation which arose by adding the dat.pl. ending to acc. **sp^he*, replacing the older form *σφι* (Ion., Hom.). This form is continued in Arcadian *σφεσιν* (SEG 37, 470.15) with *-hi* replaced by *-si(n)*, and *σφεις* (IG V 2, 6.10) with added *-s* after contraction.²¹

¹⁸ But cf. van Beek *in press*, arguing that *-μεν* was preserved longer also in South Greek, and that Proto-Greek had both **-hen* and **-men*; the choice depended on whether the paradigm had ablaut or not.

¹⁹ As for linguistic geography, features shared exclusively by non-contiguous dialects are plausibly analyzed as shared innovations stemming from an earlier period when these dialects were in direct contact.

²⁰ See Cowgill 1966 for an overview of earlier literature on the position of Mycenaean.

²¹ See the discussion in Morpurgo Davies 1992: 429–30.

Risch (e.g. 1955) claimed that there were no noticeable differences between Mycenaean and Proto-Ionic in the fourteenth or thirteenth century BCE. For this, he has been widely criticized (see Cowgill 1966). It is difficult to disprove that all characteristic innovations of Ionic-Attic (beyond general South Greek features) took place after the Mycenaean period, but Mycenaean has also undergone changes that are not paralleled in any first millennium dialect (cf. García Ramón 2016: 242–3).²²

- raising *e* > *i* before labial sounds
- palatalization of /sk/, as evidenced by the orthographic variation *a-ke-ti-ri-ja* ~ *a-ze-ti-ri-ja* /(*)askētriai/ (Méndez Dosuna 1993)
- neuter nouns in *-mo(t-)* (e.g. *pe-mo* ‘seed’) instead of *-ma(t-)*.

Several scholars have viewed these features as reflecting dialectal or sociolinguistic differences among Mycenaean scribes (“normal” vs. “special” Mycenaean, in the terms introduced by Risch 1966; monographic discussion in Hajnal 1997), but the evidence is far from clear, and it has alternatively been explained by Thompson (1996–7) as orthographic variation reflecting language change in progress.

11.3.2 *Arcado-Cypriot*

Arcadian and Cypriot are closely related South Greek dialects, but are they closer to each other than to Mycenaean or Proto-Ionic? Morpurgo Davies (1992) has shown that Proto-Arcado-Cypriot can be sensibly reconstructed. The following features are relevant:²³

- raising **en-*, *on-* > *in-*, *un-* in the preverbs/prepositions *ἐν*, *ὄν* (= Att. *ἐν*, *ὄν*)
- word-final *-o* > *-u* and diphthongization in the gen.sg. *-āo* > Arc. *-av*, Cyp. */-au/*
- analogical nom.sg. *-ης* of nouns in *-εός* (after acc. *-ην*)
- demonstrative *ὄνν* (= Ion.-Att. *ὅδε*)
- *ἀπν* and *ἐξ* governing the dative, not the genitive
- preverb/preposition /*pos*/ (Arc. *πός*, Cyp. *po-se*) instead of Ionic-Attic *πρός*
- generalization of the by-form /*kas*/ (Arc. *κας*, Cyp. *ka-se*) of the conjunction *καί*.

With the exception of some Pamphylian forms, the above isoglosses are exclusive.²⁴ Interestingly, most of the common features of Arcado-Cypriot

²² Here might also be mentioned the desyllabification of /i/ before vowels and the subsequent palatalization of velars, e.g. *su-za* /sūtʃa/ < **sūkīā* < **sūkīā* ‘fig tree’, but note that desyllabification of /i/ also occurs in Aeolic dialects.

²³ Cf. García Ramón (2010: 227–9; 2017: 78–9). This list excludes lexical choices, which mostly concern words otherwise preserved only in epic Greek, e.g. *αἶσα* ‘lot; fate’. Unlike García Ramón, I exclude the palatalization of **kʷj-* (cf. Arc. *ις*, Cyp. *si-se*, Att. *τις*) because it is not an exclusive isogloss with Arcadian, and the regular reflex of **ʔ* (Arc. has *ορ*, but the evidence from Cypriot is somewhat ambiguous).

²⁴ Another salient feature of Arcado-Cypriot, the athematic inflection of contract verbs, is shared with Aeolic (Thessalian, Lesbian). It is unclear to what extent this represents a shared

seem to be post-Mycenaean innovations: this is certain for nom.sg. *-ης* beside Myc. *-e-u* and for the syntax of *ἀπὸ* and *ἐξ*. As for the raising of *en-* and of word-final *-o*, these phenomena are not attested in Mycenaean spelling. Finally, note that Myc. has disyllabic *po-si* corresponding to /pos/, and that it may reflect either **poti* or **pti*.

Various features in which Arcadian and Cypriot diverge may be plausibly assigned to the period after 1200. Thus, the labial reflex of **k^{we}* in Cypr. *pe-i-se-i* ‘will pay’ (Att. *τεῖσει*) is the default outcome of a labiovelar, while the Arc. reflex /tʰe/ can be part of a development shared with the continuum of West Greek dialects and Ionic-Attic.

As we saw, Mycenaean has a few innovations not present in Arcadian and Cypriot, but the three dialects also share the exclusive innovation /sp^{hi}/ for /spⁱ/. Thus, both first millennium dialects reflect vernaculars spoken in the Peloponnese that diverged slightly from the administrative language written in Linear B but were closely related to it. The common innovations of Arcado-Cypriot may have come into being in the course of the thirteenth or twelfth century BCE, before the migration to Cyprus.

11.3.3 *Ionic-Attic*

Proto-Ionic can be reconstructed fairly well. Exclusive shared innovations between Attic and all Ionic dialects include:

- fronting **ā* > /æ:/
- Quantitative Metathesis (there were two rounds: one preceding and another following intervocalic *w*-loss)
- nom. and acc.pl. *ἡμεῖς*, *ἡμέας* and *ὅμεῖς*, *ὅμέας* replacing PGr. forms in **-es*, *-e* (Lesb. *ἄμμες*, *ἄμμε*)
- dat.pl. orthotonic *ἡμῖν*, *ὅμῖν* (replacing *-i(n)*, cf. Lesb. *ἄμμι*)
- athematic imperf.3pl. (and pluperfect) *-σαν*, from the sigmatic aorist, replacing **(h)an*
- 3sg. **ēs* ‘was’ (etymologically expected from **e-h₁es-t*, and attested in WGr. *ἦς*) was replaced by *ἦν* (originally 3pl. ‘were’); the latter was replaced as a 3pl. form by *ἦσαν*
- certain typical contractions (Buck 1955: 37–43), notably **ae* > Ion.-Att. *ā* (Dor. *η*).

Proto-Ionic probably underwent most of these exclusive innovations before the Ionian migrations to Asia Minor, which are conventionally dated to the mid-eleventh century.²⁵ A number of further innovations are isoglosses, due to

innovation. The athematic 3pl. secondary ending /-an/ (Arc. *ἐθέαν*, Cypr. *ka-te-ti-ja-ne*) is also found in Boeotian and is reconstructible for Proto-Ionic.

²⁵ In addition, Proto-Ionic underwent an early loss of word-initial and intervocalic **w*.

convergence, with neighboring West Greek dialects; they may have spread in the twelfth or eleventh century:

- word-internal $*_r > a\rho$ ($\rho\alpha$ in epic Greek or analogical, [van Beek 2013; 2022](#))²⁶
- the 1st compensatory lengthening and isovocalic contractions, leading to a seven-vowel system
- the 2nd compensatory lengthening
- dental outcomes of labiovelars before front vowels (cf. also Arc.)
- thematic inflection of contract verbs
- mid.3sg. $-\tau\alpha\iota \leftarrow *-toi$ (also Aeolic)
- impv.act.3pl. $-\nu\tau\omega\nu < -\nu\tau\omega + \nu$ (also in Delphic, Cretan, Thera; contrast $-\nu\tau\omega$ in most other dialects, Lesb. $-\nu\tau\omicron\nu$).

It remains uncertain as to what extent Proto-Ionic had already innovated with respect to Mycenaean-like dialects in the thirteenth century. The apparently clear distinction in the reflexes of $*_r$ (Ionic-Attic $a\rho$, Mycenaean spelled with the o -series) is difficult to use as evidence because a retention of $*_r$ in Mycenaean cannot be excluded, and the same might be true of Proto-Ionic at this date ([van Beek 2013; 2022](#)). The outcome of secondary $*_t^{(h)}i$ was Proto-Ionic $*ts$ but is spelled with the s -series in Mycenaean (e.g. *pe-de-we-sa* ‘with feet’), which may represent either /ts/ ([Crespo 1985](#)) or /ss/ ([Viredaz 1993](#)); in the latter case, Mycenaean would have innovated with respect to Proto-Ionic.

With the migrations across the Aegean, various local varieties of Ionic developed. The main division is between Western dialects (subdivided into Attic and Western Ionic) and Eastern dialects (subdivided into Central and Eastern Ionic); it includes the following characteristic innovations:

- $*ts > \sigma\sigma$ (Eastern and Central Ionic), $\tau\tau$ (Attic, Western Ionic)
- loss of $*w$ after R, s with compensatory lengthening (Eastern Ionic), or without compensatory lengthening (Attic, Western Ionic)
- $*rs > \rho\rho$ (Attic, Western Ionic)
- reversion $*\alpha\epsilon > \bar{a}$ after i, e, r (Attic, perhaps Western Ionic)
- loss of h - (Eastern Ionic)
- rhoticism, i.e. $s > r$ between vowels and word-finally (Western Ionic).

Some of these developments are shared with neighboring dialects (Boeotian, Lesbian).

11.3.4 The Unity of Aeolic and the Position of Proto-Aeolic

The need to reconstruct Proto-Aeolic has been forcefully defended by [García Ramón \(2010\)](#), reacting to the superficial treatment by [Parker \(2008\)](#).²⁷ García

²⁶ It is uncertain whether $a\rho$ or $\rho\alpha$ was the regular reflex in mainland West Greek dialects, but α as an anaptyctic vowel is certain.

²⁷ On this issue, and on the internal subgrouping of Aeolic, see also the unpublished dissertation by [Scarborough \(2016\)](#).

Ramón argues that the Aeolic dialects were linked in the twelfth century BCE not only by shared innovations but also by a number of common selections among different alternatives and common retentions.²⁸ Clear shared innovations exclusive to all three Aeolic dialects are

- $*_r > \rho o$
- labial reflexes of the labiovelars before front vowels²⁹
- $\rho i > \rho e$ (Lesb. *Δαμοκρετω* for class. *Δημοκρίτου*, Thess. *κρεννεμεν* for class. *κρίνειν*, Boeot. *τρέπεδδα* ‘table’ from **tripedza*, cf. Hsch. *τρίπεδδαν*)
- the sigmatic aorist in *-σσ-* of stems in a vowel, analogically extended from stems in *-s-*
- the perfect participle in *-οντ-*.³⁰

The change $*_r > \rho o$ has gained significance in the light of my investigation of the place of the anaptyctic vowel (van Beek 2013; 2022): the regular reflex is ρo in Aeolic dialects, but not in Mycenaean (which has either $*_r$ or $o\rho$) or Arcadian ($o\rho$). This makes $*_r > \rho o$ an exclusive innovation of all three Aeolic dialects, which may be dated to the late Mycenaean period or before.

The following features might be added:

- 3rd declension dative plural in *-εσσι*³¹
- feminine *ἴα* ‘one’ (Lesb., Thess., Boeot.) vs. *μία* (all other dialects)³²
- thematic inf. *-εμεν* (Thess. and Boeot.), but only if Lesb. *-ην* is due to Ionic influence
- temporal adverbs in *-τα* (Lesb. and Thess.), if Boeot. *-κα* is from West Greek.³³

According to Risch (1963), more fully elaborated by García Ramón (1975), there is no hard evidence for an Aeolic subgroup in the Mycenaean era. García Ramón dates the above innovations to the twelfth or even eleventh century.

²⁸ I agree with García Ramón that common choices between alternatives are also significant for subgrouping, but I disagree with his emphasis on the significance of common retentions (such as the patronymic adj. in *-ιος*, which is also preserved in Mycenaean but replaced by the gen. of the father’s name in WGr. and Ion.-Att.).

²⁹ Exceptions are the clitics *τε* < $*k^{we}$ and *τις* < $*k^{wis}$ in all three Aeolic dialects; the Perrhaebian form *κίς* may have been generalized from negated $*ou=kis$.

³⁰ For a more extensive list of features, see Méndez Dosuna 2007a. I have left aside the desyllabification $*CRiV > *CRiV$, which leads to partly different results in Thess., Boeot., and Lesb., but may still reflect an early common tendency of the three dialects (García Ramón 2010: 223–4 and 225). Hajnal (2007: 151–2) sees evidence for this change in Mycenaean and views it as an isogloss with early Aeolic.

³¹ Although *-εσσι* also occurs in some subtypes of 3rd declension stems in various West Greek dialects, it was the *only* current 3rd declension ending (excepting *s*-stems, where both *-εσσι* and *-έσσι* occur) in all three Aeolic dialects. García Ramón’s view (1975: 83–4) that it arose after the split-up of Proto-Aeolic seems unlikely to me for reasons I will discuss elsewhere.

³² The reconstruction of the PGr. form is debated: does *ἴα* reflect a reduced form $*smiā-$ > $*sīā-$ that was leveled from the oblique cases, or does it reflect a different pronominal stem? This issue does not, however, change the significance of the presence of *ἴα* in all Aeolic dialects (García Ramón 2010: 225–6).

³³ See García Ramón 2010: 232 and 2017: 43–4 on Thess. *ποτα* and *οκκε* (< $*hota=ke$).

However, a number of typical Aeolic innovations probably pre-dated the turmoil of the Dark Ages. For instance, since the Aeolic dialects were not affected by the palatalization processes of labiovelars found in West Greek, Ionic-Attic, and Arcadian, the development to labials is best seen as an earlier innovation of Proto-Aeolic. It is more likely that the differences between West Greek and Aeolic developed gradually over the course of the Mycenaean period.

Lesbian also has features not shared by Thessalian and Boeotian, including³⁴

- assibilation **ti > σι*
- preverb/preposition *πρός* (against *ποτι*)
- *o*-vocalism in *βόλλομαι* ‘want’ (against Thess. ptc. *βέλλομενος*, Boeot. *βειλομενος*)
- *εἰς*, *ἐς* (< **ens*) + acc. ‘into’ (against *ἐν* + acc.)
- thematic infinitives in *-ην* (against *-εμεν*)
- athematic infinitives in *-ν* and *-μεναι* (against *-μεν*).

These divergences are usually accounted for by assuming that the Lesbian features arose in contact with Ionic (Risch 1955). Indeed, the preverbs *πρός* and *εἰς*, *ἐς* might be borrowings from Ionic, and *βόλλομαι* might be a crossover between earlier *βέλλομαι* and Ionic *βούλομαι*. The evidence for **ti > σι*, however, is problematic: Lesbian seems to have undergone a sound change, but this would be unexpected as the result of contact since first-millennium Ionic did tolerate /ti/ again. We may therefore envisage a different scenario in which the second-millennium precursor of Lesbian took part in at least one archaic South Greek innovation (**ti > σι*) and also in the exclusive isoglosses just listed with Thessalian and Boeotian, without taking part in later exclusive South Greek innovations.³⁵ This would be compatible, for instance, with a localization of pre-Lesbian on the southeastern fringes of Thessaly, in what was certainly part of the Mycenaean realm, or even in Boeotia. In other words, Lesbian would be a bridge dialect between South Greek and Aeolic (thus already Chadwick 1956: 48).

As for Boeotian, this dialect did not undergo all the innovations shared by Thessalian and Lesbian. For this reason, García Ramón 1975 assumes that its speakers migrated into Boeotia in the mid-twelfth century, and that Thessalo-Lesbian underwent a couple of further innovations, including the characteristic Aeolic gemination (in contrast to compensatory lengthening of the vowel in most other dialects), before the Lesbian migration.

³⁴ The athematic infinitive in *-μεναι* is often included in the evidence for influence of Ionic on Lesbian: it is supposed to be a contamination of Aeol. *-μεν* and Ion. *-ναι*. However, *-μεναι* may be an archaism inherited from Proto-Greek (García Ramón 2009) or an inner-Lesbian extension of **-men*. See van Beek in press.

³⁵ Similarly, but different in the details, Finkelberg 2017. For the athematic infinitives, see van Beek in press.

11.3.5 *Doric and North West Greek Dialects as Varieties of West Greek*

West Greek dialects are characterized mainly by the absence of specific innovations of South Greek (e.g. assibilation of **ti*) and/or Aeolic (e.g. thematic inf. in *-μεν*), i.e. by retained archaisms, but they also underwent a small number of common innovations.³⁶ These pan-West Greek innovations must be projected back into the Mycenaean period: if they were later isoglosses it would be difficult to understand why Attic and Arcadian do not share them.

Innovations include:

- the so-called “Doric future” in *-σέω* (also found in all NWGr. dialects), which arose through contamination of *-σω* and the “Attic” future in *-έω*
- aorist and future stem in *-ξ-* of all verbs in *-ζω*
- the numeral *τέτορες* ‘4’, with analogical *-τ-* for **-tu-* (perhaps after **k^wet^rto-*).
- lexical: e.g. *ιαρός* instead of *ιέρως* or *ιρός*, *ἄρταμις* instead of *ἄρτεμις* (cf. also Myc. gen. *A-ti-mi-to*).

Choices between alternatives include:

- */a/* < **n* in the numerals *φίκατι* ‘20’ (also in Thess. *ικατι*, Boeot. *φικατι*, without prothetic vowel) and *-κατιοι* ‘-hundred’
- generalization of the ancient primary 1pl. ending *-μες* (SGr. and Aeol. *-μεν*)
- temporal adverbs in *-κα* (also in Boeotian); contrast SGr. *-τε*, Thess. and Lesb. *-τα*
- the anaphoric pronoun *νιν* (contrast Myc. */min/*, Ion. *μιν*)
- modal particle *κα̃*, elided *κ’* (also in Boeotian; Thess. Cyp. *κε*, Lesb. *κεν*, Arc. and Ion.-Att. *ἄν*)
- ordinals *πρῶτος* ‘first’ (also in Boeotian) vs. Att. *πρῶτος*, both from **pro-atos* (Cowgill 1970: 123 and 148), *ἕβδομος* ‘seventh’ vs. Att. *ἕβδομος*, and the cardinal *τετράκοντα* ‘forty’ vs. Att. *τετταράκοντα*.

Interestingly, West Greek dialects appear to diverge in their treatment of **r* (van Beek 2013; 2022). Cretan dialects have a regular anaptyxis before */r/*, and probably a conditioned reflex: *ap* normally, but *op* after labials. On the other hand, the dialects of Elis and Corinth (and its colony Syracuse) seem to have the regular anaptyctic vowel after */r/* (e.g. *ἔπραδες* for *ἐπαρδες* ‘you farted’ in the Syracusan poet Sophron). This would have the important consequence that Proto-West Greek retained **r* until Dorians settled on the Peloponnese and Crete in the twelfth–eleventh century BCE.

Since the nineteenth century, West Greek has been subdivided into “severe Doric” (characterized by a system with five long vowels) and “mild Doric” (seven long vowels, with */e:/* and */o:/* from contractions and the 1st compensatory lengthening, as in Ionic-Attic). In addition to this, Bartoněk (1972) pointed out the existence of “middle Doric” (seven long vowels, with */e:/* and */o:/* from

³⁶ Cf. Méndez Dosuna 2007b for a complete list including more examples, but with some different choices.

contractions, but /ε:/ and /ο:/ from the 1st compensatory lengthening). According to Bartoněk the severe Doric dialects form a distinct subgroup of West Greek, but most scholars now suppose that the various different long vowel systems of West Greek dialects took their shape in the late second / early first millennium BCE and kept developing afterwards (Méndez Dosuna 1985; Ruijgh 2007). Indeed, Elean attests yet another different system with six long vowels and its own peculiar history.

Doric and the North-Western group are best seen as deriving from a more or less undifferentiated West Greek. Except for the creation of **ens* + acc. ‘into’, which is shared with Ionic-Attic, there are no common innovations of the Doric dialects to the exclusion of NWGr. (Méndez Dosuna 1985; see Méndez Dosuna 2007b: 445 for an overview of relevant features). Moreover, due to the lacunary attestation of many North-Western dialects, it remains uncertain whether they formed a distinct branch of West Greek, or rather a convergence area.

11.3.6 *The Status of Pamphylian*

Even the few data we have for Pamphylian make it clear that the dialect cannot be assigned to one of the groups discussed above: it has, for instance, the athematic infinitive *α[φ]ιεναι* (South Greek), dative plural in *-εσσι* (Aeolic, NWGr.), *hoka* = *ὅτε*, *hapos* = *ἱερός* (West Greek only), and *φικατι* /*wīkati*/ ‘twenty’ (West Greek or Aeolic). From this, it has been concluded that Pamphylian is a mixed dialect, possibly reflecting an original Mycenaean settlement with a superposition of later West Greek and Aeolic strata (Brixhe 1976: 149; 2013: 189–203).

11.3.7 *Branching and Dating: Tentative Conclusions*

In sum, the most likely scenario is as follows (see the tentative tree in Figure 11.1). In the first centuries of the second millennium, Proto-Greek was undifferentiated, although there was no doubt some variation, as well as affinities with other Balkan languages.³⁷ Around 1700, South Greek-speaking tribes penetrated into Boeotia, Attica, and the Peloponnese, while North Greek was spoken roughly in Thessaly, parts of Central Greece, and further North and West (up to Epirus, and perhaps also Macedonia). During the early Mycenaean period, South Greek diverged by the assibilation of **ti*, the simplification of word-internal **ts* and **ss*, and a number of morphological innovations.

³⁷ Scholars often date the immigration into the Peloponnese to the end of the third millennium, but I would prefer a later date coinciding with the beginning of Late Helladic, in the seventeenth century BCE (cf. Hajnal 2005). This would fit the linguistic data best, as reconstructible differences between South Greek and North Greek in the late Mycenaean period are relatively small.

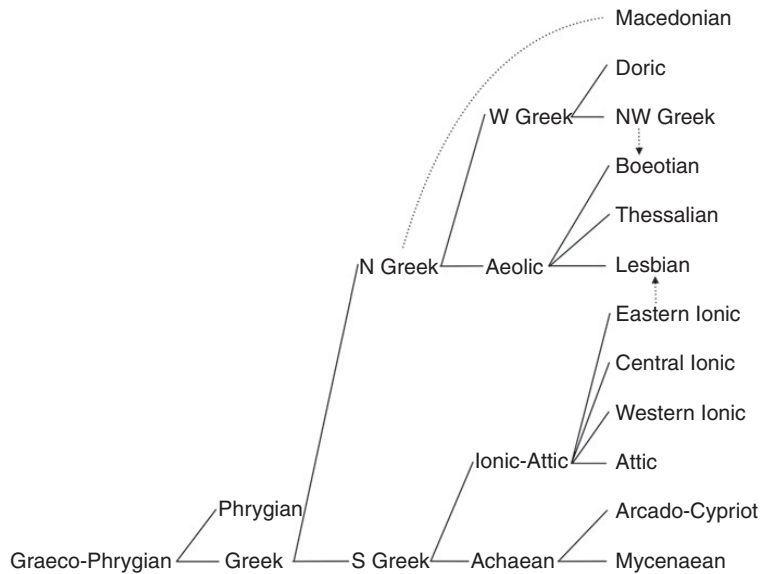


Figure 11.1 The Greek dialects

At some point, probably still in the Mycenaean period, Proto-Aeolic developed as a result of changes such as $*r > \rho o$, labial reflexes of all remaining labiovelars, and the creation of 3rd decl. dat.pl. $-\varepsilon\sigma\sigma\iota$. Proto-Aeolic can be reconstructed if the South Greek features of Lesbian and the West Greek features of Boeotian can be ascribed to contact with Ionic and West Greek, respectively, in the late Dark Ages. Alternatively, the precursors of Lesbian and Boeotian in the Mycenaean period may have been bridge dialects linking Thessalian with South Greek and West Greek, respectively.

In the thirteenth–twelfth century BCE, then, there were (at least) three larger dialect areas: South Greek on the Peloponnese and in Attica and Boeotia; Aeolic in Thessaly, and West Greek in North-Western regions. Moreover, in the same period Proto-Ionic also started to diverge from Mycenaean-like dialects (Proto-Arcado-Cypriot). We are in the dark, however, about the dialects spoken in Central Greece, and not all dialects spoken in this period need have survived.

The traditional concept of Dorian migrations in the twelfth and eleventh centuries is still the best way to explain the isolated position of Arcadian and the specific institutions shared by various Dorian states. Many defining characteristics of the first-millennium dialects (including isoglosses shared between Proto-Ionic and West Greek) took shape in the Dark Ages through convergent

developments; this means that the situation in the second millennium may have been quite different (cf. the discussion about the position of Aeolic), and many specific details cannot be recovered.

11.4 The Relationship of Greek to the Other Branches

11.4.1 Greek and Macedonian

Macedonian is known from various Greek-like personal names, some glosses in Hesychius, and probably from a curse tablet found at Pella, containing an unknown form of Greek resembling NWGr. dialects (*SEG* 43.434, c. 380–350 BCE, [Hatzopoulos 2007](#)). To this might be added an oracular consultation on a lead tablet found at Dodona (Méndez [Dosuna 2012](#): 144–5). The Pella curse tablet shares some typical features with NWGr. dialects: apocope in the preverb *κατ-*, dat. pron. *ἐμίν* vs. *ἐμοί*, and a temporal adverb in *-κα*. On the other hand, scholars have traditionally viewed Macedonian as a separate language closely related to Thracian and Phrygian on account of reflexes of the “voiced aspirates” written <β δ γ> (e.g. *Βουλομαγα* = *Φυλλομάχη*). However, this does not explain e.g. the reflex of **g^h*- in the name *Κεβαλιος* (cf. Gr. *κεφαλή*): if Macedonian had a Thracio-Phrygian-like development, one would expect **Γεβαλιος*. Moreover, since there is also evidence that voiceless stops were voiced between vowels and in contact with sonorants (e.g. *διγαία* = Att. *δικαία*, *Δρεβέλαος* = Att. *Τρεφέλεως*), it is proposed (cf. Méndez [Dosuna 2012](#)) that <β δ γ> may represent both voiced fricatives (from **p^h t^h k^h*) and normal voiced stops (**p t k*); finally, *Κεβαλιος* presupposes that Macedonian took part in Grassmann’s Law. If this is correct, Macedonian started off as a NWGr. dialect which subsequently underwent its proper *Lautverschiebung* in the stops. Caution is obviously necessary in view of the limited evidence.

11.4.2 Greek and Phrygian

Greek is clearly more closely related to Phrygian than to any of the main branches of Indo-European: there are shared phonological, morphological and lexical innovations.³⁸ This close correspondence is all the more remarkable given the fragmentary attestation of Phrygian. The view that Phrygian and Armenian are especially closely related, already expressed in ancient authors, is not based on compelling evidence (cf. [Obrador-Cursach 2019](#): 240–2; *contra* [Lamberterie 2013](#)).

³⁸ See [Neumann 1988](#), [Lamberterie 2013](#) and [Obrador-Cursach 2019](#) on Graeco-Phrygian, and [Ligorio & Lubotsky 2018](#) for a recent encyclopedic treatment of Phrygian.

Phrygian shares phonological innovations such as the following with Greek:

- a threefold reflex of PIE **CRHC* is proven by MPhr. *γλουρεος* ‘golden’ (cf. *γλούρεα*: *χρύσεια*. *Φρύγες* < *καὶ* > *γλουρός*: *χρυσός*, Hsch. γ 659), corresponding to Greek *χλωρός* ‘bay, pale; green’ < PIE **ǵʰh₃-ró-*; this development is not shared with any other Indo-European language
 - a threefold reflex of word-initial **HC-*, cf. NPhr. *αἶψα* < **h₂nēr* (Gr. *ἀνήρ*), OPhr. *onoman* (Gr. *ὄνομα*)³⁹
 - triple reflex of PIE **CHC*: Phr. *-μενος* < **-mh₁nos*, as in Greek
 - lenition of prevocalic **s*, word-initially (NPhr. *εγεδον* = Gr. *ἐχέσθω* < **seǵʰ-*) and after a vowel (NPhr. *δεως* = Gr. *θεοῖς* < **dʰh₁s-ó-*), as well as in **sw-*
 - loss of word-final occlusives: 3sg. impv. *-tov* = Gr. *-tōd*.
- Note that Phrygian is a *centum* language: cf. OPhr. *egeseti*, NPhr. *εγεδον* < **seǵʰ-e/o-*; MPhr. *γλουρεος* < PIE **ǵʰh₃-ró-* plus **-eǵios*. Other phonological innovations led to differences with Greek, but none of them has to be early:
- the labiovelars were merged with the pure velars and palato-velars: NPhr. *κναικαν* = Gr. *γυναῖκα*
 - the PIE voiced obstruents developed into voiceless stops (Lubotsky 2004): acc. *Τιαν* = *Ζῆν(α)*, gen. *Τιος* = *Διός*, dat./instr. *Τι(ε)* = *Δίι*, *Δίι*, as well as acc. *κναικαν* ‘wife’ = Gr. *γυναῖκα*.

The following morphological isoglosses are relevant:

- OPhr. (probably 3sg. opt.) *kakoiou*, *kakuioy*, probably a counterpart to Greek *κακῶ* ‘maltreat’ with preserved intervocalic *yod*; both the type of factitive formation and the lexeme are exclusive to Phrygian and Greek
- OPhr. *avtos*, an exclusive isogloss with Gr. *αὐτός* ‘self’, cf. (38) above; the combination OPhr. *venavtun*, with secondary *-n-*, neatly matches Gr. *ἐαυτόν* ‘himself’ < **swe auton*
- the suffix **-ēu-* in Greek masculine nouns in *-εύς* seems to be matched by (apparently thematized) OPhr. *-avo-*
- NPhr. 3sg. *εγεδον*, probably a middle imperative, is paralleled by Gr. *-σθω* (possibly a common innovation, Ligorio & Lubotsky 2018: 1828)
- the middle perfect ptc. in *-μενος* < **-mh₁nos* (formed in an identical way in Greek).

Phrygian preserves several morphological archaisms that Proto-Greek lost. The 3pl. perfect ending **-ēr* is probably continued in NPhr. *δακαρεν* ‘they established’ (**-ēr* plus **-ent*). On the whole, however, the Phrygian verb displays many innovations, even if most details are still unclear.

³⁹ The Armenian reflexes of these words (*ayr* ‘husband’, *anun* ‘name’) also have “prothetic vowels”; this is often interpreted as a common development of “Balkanindogermanisch” (cf. Hajnal 2003), but the laryngeals developed differently in Armenian in other environments, whereas there are no discernable differences between Greek and Phrygian.

Lexically, the following items are important:

- Phryg. *knaikan* ‘woman, wife’ beside Gr. *γυναικα*, reflecting PIE **g^wen-h₂*, **g^wn-eh₂-* with an additional suffix *-ik-* (or *-i-k-*: cf. Armenian pl. *kanai-k* ‘women’ without the *k*-suffix)
- Gr. *ὄνομα* ‘name’ and Phryg. *onoman* ‘id.’ with a zero grade root (also attested elsewhere, but contrast Latin *nōmen*, Vedic *nāman-*, Armenian *anun* < **o/anōmn*)⁴⁰
- Phr. *δεως* (instr.pl.) and Gr. *θεός* reflect PIE **d^hh₁s-ó-* ‘god’, while most other languages have a reflex of **deiyó-*
- NPhr. *υψοδαν*, if reflecting an adverb **ups-o-d^hh₁* ‘above’, forms a near-precise match with Gr. *ὑψόθεν* ‘on high; from above’ (Lubotsky 1993).

Notwithstanding the fragmentary attestation of Macedonian and Phrygian, it seems likely that their ancestors formed a linguistic unity with (pre-)Proto-Greek in the late third and early second millennium BCE, presumably somewhere on the southern Balkans (Macedonia, Thracia), before Hellenes penetrated into Thessaly and further south. The relationship to other Balkan languages remains quite uncertain. Hajnal (2003) collects some possible evidence for prehistoric contacts between Ancient Balkan languages, including the appurtenance suffix *-ejo-* (attested in Greek, probably in Phrygian *kubeleya*, possibly in Venetic and Messapic, but not elsewhere) and the innovative dat.-loc. ending *-si* (probably found in Albanian *-sh*),⁴¹ but there is not enough evidence for drawing solid conclusions.

11.4.3 Greek and Armenian

The possibility of a closer relation between early forms of Greek and Armenian has attracted scholarly attention since the works of Meillet and Pedersen. In more recent times, a genealogical connection has been pleaded for by Olsen & Thorsø (Chapter 12) and Lamberterie (1997; 2013). Skepticism has been voiced by Clackson (1994) and, recently, Kim (2018). Indeed, there are no phonological isoglosses that must be distinctive innovations shared exclusively by Greek and Armenian, and what are probably the earliest phonological innovations of Armenian are generally not matched by Greek counterparts. Furthermore, shared morphological innovations cannot be demonstrated (Clackson 1994: 60–87).

Having said this, certain lexical isoglosses remain suggestive, especially those that combine semantic and morphological developments. For an overview of lexical correspondences between Armenian, Greek, and Indo-Iranian,

⁴⁰ Phrygian *onoman* renders highly unlikely the idea that the initial vowel of Laconian *Εννομακρατιδής* directly reflects PIE **h₁ǵh₃-m̥n* and that *ὄνομα* arose by vowel assimilation (cf. Lamberterie 2013: 34 with references). The root of ‘name’ must therefore be PIE **h₃neh₃-*.

⁴¹ Note that the existence of a Phrygian dative in *-ωσι* (admitted by Hajnal 2003) is uncertain.

see [Martirosyan \(2013\)](#) and Olsen & Thorsø ([Chapter 12](#)), though part of the material consists of shared retentions and independent borrowings. The following examples are among the strongest:

- Gr. ἡμᾶρ < **ām̥r* ~ Arm. *awr* ‘day’ < **āmōr* or **ām̥r* (cf. [Kim 2018](#): 252), a (near-)perfect word-equation: this isogloss of core vocabulary is exclusive to Armenian and Greek, but Ved. *āhar* (gen. *āhnas*) and Av. *aīiarə* ‘day’ look suspiciously similar to each other and to the Graeco-Armenian word. It cannot be ruled out that **ām̥r* reflects an archaism of PIE ([Clackson 1994](#): 97; [Pinault 2017](#)).
- The full grade root of *δηρός* and Arm. *erkar* ‘long’ < **duāró-* is certainly an innovation of both branches, whether it is the phonological outcome of **duh₂r-ó-* or an analogical reshaping **duh₂r-ó-* after the adverb **duh₂m* (cf. Gr. *δήν*, Arm. *erkayn* < **duān-jo-*, Old Hittite *tūuaz* ‘from afar’).
- The reduplicated aor. **ar-ar-e/o-* (Arm. *arari* ‘made’, Gr. ἥραρον ‘fixed’) looks like an innovation: full reduplication with vowel-initial roots was productive in Greek, but not in PIE or Armenian; on possible reconstructions of the pre-form, see [Willi 2018](#): 80–2, who prefers the scenario that an original **h₂e-h₂r-e/o-* (> **āre/o-*) was restored as **h₂r-h₂r-e/o-* before the laryngeals were eliminated.
- Gr. *θερμός* and Arm. *ǰerm* ‘warm’ < **g^{wh}er-mó-*, with *e*-grade root as opposed to the *o*-grade in most other branches (Lat. *formus*, Eng. *warm*). The innovation seems due to influence of the precursor of *θερμαί* ‘become hot’ (rather than that of the nominal form *θερος* ‘heat, summer’, as per [Lamberterie 2013](#): 20), cf. also the noun Alb. *zjarm* ‘fire’ and perhaps the Phrygian toponym *Γέρμη*, *Germe*.
- **mṛtó-* ‘mortal, man’: this combination of form and meaning occurs only in Gr. *βροτός* and Arm. *mard* ([Lamberterie 1997](#)); in Indo-Iranian **mṛtá-* means ‘dead’, as expected.
- The root **h₃b^hel-* underlying Gr. *ὀφέλλω* ‘to be useful, cause to grow’, *ὄφελος* ‘benefit’ reappears in Arm. *y-awelum* ‘to add to’, aor. *y-aweli*, adv. *aweli* ‘more’; the homonymous root of *ὀφέλλω* ‘sweep’, *ὄφελμα* ‘broom’ (both only in Hipponax) recurs in Arm. *awel* ‘broom’. The root is not attested in other branches. [Clackson \(1994](#): 157) argues that the meaning ‘sweep’ is original; Greek and Armenian would both preserve the derived meaning ‘increase’, too.
- Gr. *ψεῦδομαι* ‘deceive, lie’, *ψεῦδος* ‘lie’ with Arm. *sowt*, gen. *stoy* ‘false’: the root is not attested elsewhere.

Whether such examples are sufficient for reconstructing a Graeco-Armenian node remains uncertain, as the lack of ascertained common morphological innovations is worrying. The strongest cases by comparison are

- Arm. 1sg. middle *-m* may match Greek *-μαι*, but Albanian and Tocharian also have an *m*-ending, so independent innovations cannot be excluded.

- The parallels in the formation of nasal present stems in both branches seem suggestive, but they are not numerous and are often inexact. Since double infix presents of the type *λαμβάνω* are productive in Greek beside thematic aorists, they need not be genetically related to Armenian presents in *-anem*. Thus, Arm. *lk'anem* 'leave' has been compared to *λαμβάνω*, but the latter is not attested in Homer and may be a productive creation based on *ἐλιπον* (replacing *λείπω*), while the idea that Arm. *lk'anem* < **lik^w-ane/o-* arose from **link^wη-* by dissimilation remains conjectural.
- Gr. *οὐ*, *οὐκ* 'not' and Arm. *oč'* have been derived from *(*ne*) ... *h₂oġu k^wid* by Cowgill 1960. However, Clackson (2005: 155–6) argued that *oč'* originally meant 'no one' and goes back to *o-* (as in *ok'* 'anyone' and *omn* 'someone') plus an older negation **č'* (as in *č'ik'* 'nothing') that developed from *(*ne*) ... *k^wid*. Since the loss of **ne* (e.g. French *pas*, *rien*, etc.) and the development from indefinite 'no one' to 'not' (e.g. Eng. *not*, Germ. *nicht* < **ni wihti* 'nothing') are both easily paralleled, the value of this isogloss is limited.

Finally, a number of alleged exclusive isoglosses are less strong than they seem:

- Gr. *κίων* 'pillar' matches Arm. *siwn* 'id.' < PIE **kiHuōn*, but the formation may have been present in Indo-Iranian, too (cf. Martirosyan 2013: 119, following Lubotsky).
- Arm. *merj* 'near' and Gr. *μέχρι* 'as long as, until, etc.' may reflect the same formation **me-ǵ^hsr-i* 'at hand', but the semantic divergence between *merj* and *μέχρι* is considerable (cf. Clackson 1994: 150–1), and **me-ǵ^hsr-i* would have to be an archaism of PIE.
- Arm. *artewan*, gen.pl. *-ac'* 'eyebrow' yields an exact correspondence to Gr. *δρεπάνη* 'sickle', with a metaphorical meaning of the body part in Armenian. However, the fact that *δρεπάνη* looks like an instrument noun productively derived from *δρέπω* 'pluck' casts doubt on its antiquity. Could the word be a borrowing from Anatolian Greek into pre-Armenian (cf. Clackson 1994: 190)?
- Gr. *πρέπω* 'be conspicuous' (Hom.) with Arm. *erewim* 'appear' might be an exclusive lexical isogloss if the pre-form is **prep-*, though OIr. *richt* 'form, species' might derive from **pr̥ptó-*. Alternatively, if Ved. instr. *kṛpā* 'beauty' is related, the root would be **k^wrep-*, and the verb a retained archaism.
- The word for 'goat' is Arm. *ayc* (*i*-stem) and Gr. *αἴζ*, *αἰγός*. Both derive from **aig-* or **h₂eiǵ-*; the latter is to be preferred if Av. *izaēna-* 'of leather' contains an ablauting root variant. A PIE word for 'goat' is difficult to reconstruct, and probably a borrowing.
- The meaning 'laugh' of the root **ǵelh₂-* (Gr. *γελάω* 'laugh', *γέλως* 'laughter'; Arm. *calr* 'id.', gen. *calow*) is a shared innovation. If the root of Lat. *gelidus* 'cold', *gelu* 'ice' is related (suggested by Clackson 1994: 131, positing

a development ‘shine’ > ‘ice’), the root itself is an archaism. In this case, the lexical development to ‘smile, laugh’ may have taken place in PIE, with Gr. preserving the older root meaning ‘resplendent/icy calm’ beside it.

- The formations of Arm. *nor* ‘young’ < **neuo-ro-* and *dalar* ‘green’ < **d^hIH-ro-* are not identical with Gr. *νεαρός* ‘juvenile, fresh’ and *θαλερός* ‘abundant, fertile’, respectively (note the different meaning of the latter). A relatively recent derivation of *νεαρός* and *θαλερός* within Greek is more likely (van Beek 2021b).

To conclude, I fully concur with Kim’s words (2018: 263):

[T]he list of linguistic innovations exclusively shared by Greek and Armenian is overwhelmingly composed of lexical items. Furthermore, most of these involve general root cognations, not full word equations allowing for reconstruction of an intermediate preform, which raises the possibility that they are either (partial) independent creations or even borrowings from a third language. In this respect, the relationship between Greek and Armenian differs greatly from that of Indo-Aryan and Iranian, or Baltic and Slavic, where it is possible to reconstruct dozens of distinct lexical preforms for Proto-Indo-Iranian and Proto-Balto-Slavic, respectively.

11.4.4 *Greek and Albanian*

I cannot discuss the evidence for common innovations of Greek and Albanian in any detail here; for a list of potential cases, see Chapter 12, where Hyllested and Joseph adduce some interesting examples, such as the element **kīā-* (contained in both Alb. *sot* ‘today’ and Greek *τίμερον* ‘id.’). However, a number of Greek innovations adduced there can or must in my view be dated later than Proto-Greek. I am not convinced of a close genetic relation between Greek and Albanian.

11.5 The Position of Greek

The further position of Graeco-Phrygian in the family tree is not easy to determine. It is customary, and indeed plausible, to include Greek in a putative group of “Central” Indo-European languages (including Armenian, Indo-Iranian, and probably other *satem* languages) that remained in the homeland after the departure of Anatolian, Tocharian, Italo-Celtic, and perhaps Germanic. However, as with Graeco-Armenian (Section 11.4.3), the strongest affinities with Indo-Iranian are lexical (Euler 1979). Further qualitative linguistic evidence for “Graeco-Aryan” is meagre. In the phonological domain there are no demonstrable shared innovations (cf. Section 11.2 on the syllabic nasals), and those Greek innovations that are difficult to duplicate are without parallels in the other branches (e.g. the voiceless aspirate stop series, the double outcome of initial *yod*). In verbal morphology, Greek and Indo-Iranian

preserved more archaisms than most branches, partly because of their early attestation: these include the distinctions between active and middle voice, three different “tense-aspect” stems (present, aorist, and perfect), subjunctive and optative, and so on.

It is often asserted that certain similarities between the verbal systems of Greek and Indo-Iranian are common innovations. Thus, the augment, the middle perfect, and the pluperfect are ascribed to this late stage of PIE. However, the augment may well be an archaic feature. Given that Indo-Iranian uses the stative ending **-o* in the middle perfect while Greek uses middle **-to*, an independent innovation of this formation is possible. This leaves us with the creation of primary middle endings in *-i*, which might be shared with Indo-Iranian and Germanic, and the use of the originally contrastive suffix **-tero-* in comparative adjectives (shared only with Indo-Iranian).

In sum, from a qualitative angle it remains uncertain when exactly Greek (Graeco-Phrygian) branched off from Nuclear PIE. There are no indications for an early separation (which would require demonstrating a common innovation of most other branches that Proto-Greek did not undergo). A relatively late departure therefore seems likely, but the evidence for this is mainly lexical.

References

- Bartoněk, Antonín. 1972. *Classification of the West Greek dialects at the time about 350 B.C.* Amsterdam: Hakkert.
- van Beek, Lucien. 2013. *The development of the PIE syllabic liquids in Greek*. PhD thesis, Leiden University.
- van Beek, Lucien. 2019. A look into the Indo-European bedroom: Vedic *yóni-* and Greek *ἐώνη*. *Historische Sprachforschung* 132: 4–34.
- van Beek, Lucien. 2021a. Accentuation versus syllable structure: What conditioned the disyllabic reflex of PIE **CRHC* in Greek? Paper presented at the Oxford Workshop on Indo-European Accentuation, 15–16 July 2021, online/Wolfson College.
- van Beek, Lucien. 2021b. Les adjectifs en *-ερός*, *-αρός* et *-ηρός* chez Homère et ultérieurement: Origines et diffusion. In Alain Blanc & Isabelle Boehm (eds.), *Dérivation nominale et innovations dans les langues indo-européennes anciennes*, 161–82. Lyon: MOM.
- van Beek, Lucien. 2022. *The reflexes of syllabic liquids in Ancient Greek: Linguistic prehistory of the Greek dialects and Homeric Kunstsprache*. Leiden: Brill.
- van Beek, Lucien. In press. Athematic infinitives in Lesbian, Homer and other Greek dialects: Innovations, archaisms or contact-induced borrowings? In Alcorac Alonso Déniz et al. (eds.), *Contacts linguistiques en Grèce ancienne: diachronie et synchronie (CoLiGA)*. Lyon: MOM.
- Beekes, Robert. 2014. *Pre-Greek: Phonology, morphology, lexicon*. Leiden: Brill.
- Bowie, Angus. 1981. *The poetic dialect of Sappho and Alcaeus*. New York: Arno.
- Brixhe, Claude. 1976. *Le dialecte grec de Pamphylie: Documents et grammaire*. Paris: Adrien-Maisonneuve.

- Brixhe, Claude. 2013. La Pamphylie. Peuplement et dialecte: 40 ans de recherche. *Kadmos* 52. 169–205.
- Buck, Carl. 1955. *The Greek dialects*. Chicago, IL: University of Chicago Press.
- Chadwick, John. 1956. The Greek dialects and Greek prehistory. *Greece and Rome* 3. 38–50.
- Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.
- Clackson, James. 2005. Review of Frederik Kortlandt, *Armeniaca: Comparative notes*, Ann Arbor, MI: Caravan Books, 2003. *Annual of Armenian Linguistics* 24–5. 153–8.
- Clackson, James. 2007. The genesis of Greek. In A.-F. Christidis (ed.), *A history of Ancient Greek: From the beginnings to Late Antiquity*, 185–92. Cambridge: Cambridge University Press.
- Crespo, Emilio. 1985. Palatal stops in Greek: Reconstruction or Mycenaean evidence? *Minos* 19. 91–104.
- Cowgill, Warren. 1960. Greek *ou* and Armenian *oč'*. *Language* 36. 347–50.
- Cowgill, Warren. 1966. Ancient Greek dialectology in the light of Mycenaean. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 77–95. Berkeley: University of California Press.
- Cowgill, Warren. 1970. Italic and Celtic superlatives and the dialects of Indo-European. In George Cardona, Henry Hoenigswald & Alfred Senn (eds.), *Indo-European and Indo-Europeans*, 113–53. Philadelphia: University of Pennsylvania Press.
- Driessen, Jan. 2008. Chronology of the Linear B texts. In Yves Duhoux & Anna Morpurgo Davies (eds.), *A companion to Linear B: Mycenaean Greek texts and their world*. Vol. 1, 69–79. Louvain-la-Neuve: Peeters.
- Euler, Wolfram. 1979. *Indoiranisch-griechische Gemeinsamkeiten der Nominalbildung und deren indogermanische Grundlagen*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Finkelberg, Margalit. 2017. Lesbian and Mainland Greece. In Georgios K. Giannakis, Emilio Crespo, & Panagiotis Filos (eds.), *Studies in Ancient Greek dialects: From Central Greece to the Black Sea*, 447–56. Berlin: de Gruyter.
- García Ramón, José Luis. 1975. *Les origines postmycéniennes du groupe dialectal éolien: Étude linguistique*. Salamanca: Universidad de Salamanca.
- García Ramón, José Luis. 1999. Griechisch Ζητήρ· Ζεὺς ἐν Κύπρῳ, vedisch *yātar-* und die Vertretung von **j* im Griechischen. In Heiner Eichner, Hans Christian Luschützky & Velizar Sadoski (eds.), *Compositiones Indogermanicae: Gedenkschrift für Jochem Schindler (1944–1994)*, 77–96. Prague: Enigma.
- García Ramón, José Luis. 2009. Formal correspondences, different functions: On the reconstruction of inflectional categories of Indo-European. In Vit Bubenik, John Hewson, & Sarah Rose (eds.), *Grammatical change in Indo-European languages*, 237–50. Amsterdam: Benjamins.
- García Ramón, José Luis. 2010. On the genetic classification of the Ancient Greek dialects: Comparative reconstruction versus hypercriticism and atomism at work. In *Studies in Greek linguistics*, 219–236. Thessaloniki: Aristotéleio Panepistímio Thessaloníkis, Institoúto Neollinikón Spoudón.
- García-Ramón, José Luis. 2016. Il Greco miceneo. In Maurizio del Freo & Massimo Perna (eds.), *Manuale di epigrafia micenea: Introduzione allo studio dei testi in lineare B*, 211–44. Padua: librariauniversitaria.it.

- García Ramón, José Luis. 2017. Ancient Greek dialectology: Old and new questions, recent developments. In Georgios K. Giannakis, Emilio Crespo, & Panagiotis Filos (eds.), *Studies in Ancient Greek dialects: From Central Greece to the Black Sea*, 29–106. Berlin: de Gruyter.
- Garrett, Andrew. 2006. Convergence in the formation of Indo-European subgroups: Phylogeny and chronology. In Peter Forster & Colin Renfrew (eds.), *Phylogenetic methods and the prehistory of languages*, 139–51. Cambridge: McDonald Institute for Archaeological Research.
- Hackstein, Olav. 2010. The Greek of epic. In Egbert J. Bakker (ed.), *A companion to the Ancient Greek language*, 401–23. Chichester: Blackwell.
- Hajnal, Ivo. 1997. *Sprachschichten des mykenischen Griechisch*. Salamanca: Universidad de Salamanca.
- Hajnal, Ivo. 2003. Methodische Vorbemerkungen zu einer Palaeolinguistik des Balkanraums. In Alfred Bammesberger & Theo Vennemann (eds.), *Languages in prehistoric Europe*, 117–45. Heidelberg: Winter.
- Hajnal, Ivo. 2005. Das Frühgriechische zwischen Balkan und Ägäis – Einheit oder Vielfalt? In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 185–214. Wiesbaden: Reichert.
- Hajnal, Ivo. 2007. Die Vorgeschichte der griechischen Dialekte: ein methodischer Rück- und Ausblick. In Ivo Hajnal (ed.), *Die altgriechischen Dialekte: Wesen und Werden*, 131–56. Innsbruck: Institut der Sprachen und Literaturen der Universität Innsbruck.
- Hatzopoulos, Miltiades B. 2007. La position dialectale du macédonien à la lumière des découvertes épigraphiques récentes. In Ivo Hajnal (ed.), *Die altgriechischen Dialekte: Wesen und Werden*, 157–76. Innsbruck: Institut der Sprachen und Literaturen der Universität Innsbruck.
- IG V 2 = Friedrich Hiller von Gaertringen (ed.). 1913. *Inscriptiones Graecae*. Vol. 5(2). *Inscriptiones Arcadiae*. Berlin: de Gruyter.
- Kim, Ronald I. 2018. Greco-Armenian: The persistence of a myth. *Indogermanische Forschungen* 123. 247–71.
- Kortlandt, Frederik. 1981. 1st sg middle $*-H_2$. *Indogermanische Forschungen* 86. 123–36.
- Lamberterie, Charles de. 1997. Review of Clackson 1994. *Kratylos* 42. 71–8.
- Lamberterie, Charles de. 2013. Grec, phrygien, arménien: des anciens aux modernes. *Journal des Savants* 2013(1). 3–69.
- Ligorio, Orsat & Alexander Lubotsky. 2018. Phrygian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 1816–31. Berlin: De Gruyter Mouton.
- Lubotsky, Alexander. 1993. New Phrygian $\psi\phi\omicron\delta\alpha\nu$. *Kadmos* 32. 127–34.
- Lubotsky, Alexander. 2004. The Phrygian Zeus and the problem of the “Lautverschiebung”. *Historische Sprachforschung* 117. 229–37.
- Martirosyan, Hratch. 2013. The place of Armenian in the Indo-European language family: The relationship with Greek and Indo-Iranian. *Journal of Language Relationship* 10. 85–138.
- Méndez Dosuna, Julián. 1985. *Los dialectos dorios del noroeste: Gramática y estudio dialectal*. Salamanca: Universidad de Salamanca.
- Méndez Dosuna, Julián. 1993. A note on Myc. *a-ze-ti-ri-ja*, Att. $\sigma\beta\acute{\epsilon}\nu\nu\mu\iota$, and palatalization. *Die Sprache* 35. 208–20.

- Méndez Dosuna, Julián. 2007a. The Aeolic dialects. In A.-F. Christidis (ed.), *A history of Ancient Greek: From the beginnings to Late Antiquity*, 460–74. Cambridge: Cambridge University Press.
- Méndez Dosuna, Julián. 2007b. The Doric dialects. In A.-F. Christidis (ed.), *A history of Ancient Greek: From the beginnings to Late Antiquity*, 444–59. Cambridge: Cambridge University Press.
- Méndez Dosuna, Julián. 2012. Ancient Macedonian as a Greek dialect: A critical survey on recent work. In G. K. Giannakis (ed.), *Ancient Macedonia: Language, history, culture*, 133–45. Thessaloniki: Kéntro Ellinikis Glóssas.
- Morpurgo Davies, Anna. 1992. Mycenaean, Arcadian, Cyprian and some questions of method in dialectology. In J.-P. Olivier (ed.), *Mykenaika: Actes du IX^e Colloque international sur les textes mycéniens et égéens organisé par le Centre de l'Antiquité Grecque et Romaine de la Fondation Hellénique des Recherches Scientifiques et l'École française d'Athènes (Athènes, 2–6 octobre 1990)*, 415–31. Paris: de Boccard.
- Neumann, Günter. 1988. *Phrygisch und Griechisch*. Vienna: Österreichische Akademie der Wissenschaften.
- Obrador-Cursach, Bartomeu. 2019. On the place of Phrygian among the Indo-European languages. *Journal of Language Relationships* 17, 233–45.
- Parker, Holt. 2008. The linguistic case for the Aiolian migration reconsidered. *Hesperia* 77, 431–64.
- Peters, Martin. 1980. *Untersuchungen zur Vertretung der indogermanischen Laryngale im Griechischen*. Vienna: Österreichische Akademie der Wissenschaften.
- Pinault, Georges-Jean. 2017. In the heat of the day. In Ivo Hajnal, Daniel Kölligan & Katharina Zipser (eds.), *Miscellanea Indogermanica: Festschrift für José Luis García Ramón zum 65. Geburtstag*, 651–69. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Porzig, Walter. 1954. Sprachgeographische Untersuchungen zu den altgriechischen Dialekten. *Indogermanische Forschungen* 61, 147–69.
- Risch, Ernst. 1955. Die Gliederung der griechischen Dialekte in neuer Sicht. *Museum Helveticum* 12, 61–76.
- Risch, Ernst. 1963. Il problema dell'unità linguistica greca. In B. E. Vidos, *Le proto-lingue*, 91–109. Milan: Paideia.
- Risch, Ernst. 1966. Les différences dialectales dans le mycénien. In L. R. Palmer & John Chadwick (eds.), *Proceedings of the Cambridge Colloquium on Mycenaean Studies*, 150–7. Cambridge: Cambridge University Press.
- Rix, Helmut. 1992. *Historische Grammatik des Griechischen*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Ruijgh, Cornelis J. 2007. L'évolution des dialectes doriens jusqu'à la koina dorient: Le système des voyelles longues et la formation du futur. In Ivo Hajnal (ed.), *Die altgriechischen Dialekte: Wesen und Werden*, 393–447. Innsbruck: Institut der Sprachen und Literaturen der Universität Innsbruck.
- Scarborough, Matthew. 2016. *The Aeolic dialects of Ancient Greek: A study in historical dialectology and linguistic classification*. DPhil dissertation, University of Cambridge.
- Schrijver, Peter. 1999. Griechisch ᾗδῃ 'er wußte'. *Historische Sprachforschung* 112, 264–72.
- SEG 37 = H. W. Pleket & R. S. Stroud (eds.). 1990. *Supplementum epigraphicum Graecum* 37. Amsterdam: Gieben.

- SEG 43 = H. W. Pleket, R. S. Stroud & Johan Strubbe (eds.). 1996. *Supplementum epigraphicum Graecum* 43. Amsterdam: Gieben.
- Thompson, Rupert. 1996–7. Dialects in Mycenaean and Mycenaean among the dialects. *Minos* 31–33. 313–33.
- Tribulato, Olga. 2019. Lingue letterarie e dialetti nell'esegesi antica. In Andreas Willi (ed.), *Formes et fonctions des langues littéraires en Grèce ancienne*, 359–87. Geneva: Fondation Hardt.
- Tucker, Elizabeth. 1990. *The creation of morphological regularity*. Göttingen: Vandenhoeck und Ruprecht.
- Vine, Brent. 1999. On “Cowgill’s Law” in Greek. In Heiner Eichner, Hans Christian Luschützky & Velizar Sadoski (eds.), *Compositiones Indogermanicae: Gedenkschrift für Jochem Schindler (1944–1994)*, 555–99. Prague: Enigma.
- Viredaz, Rémy. 1993. Palatalisations grecques: Chronologie et classification des traitements. In Emilio Crespo, José Luis García Ramón, & Araceli Striano (eds.), *Dialectologica greca*, 331–7. Madrid: Ediciones de la universidad autónoma.
- Waal, Willemijn. 2018. On the “Phoenician letters”: The case for an early transmission of the Greek alphabet from an archaeological, epigraphic and linguistic perspective. *Aegean Studies* 1. 83–125.
- Willi, Andreas. 2018. *Origins of the Greek verb*. Cambridge: Cambridge University Press.

12 Armenian

Birgit Anette Olsen & Rasmus Thorsø

12.1 Introduction

The attestation of the Armenian language begins in the early fifth century where, according to tradition, the clergyman Mesrop Maštoc' invented the Armenian script for the purpose of translating the Bible. This century marks the initial period, the “golden age” (*oskedar*) of Classical Armenian or *grabar* (written language). Besides the Bible, the earliest texts consist of translations from Greek and Syriac, but also a number of original works. These include for example Eznik's “Refutation of the sects”, Koriwn's “Life of Maštoc'” and, a little later, the historical works by Agat'angelos, P'awstos Bowzand, Łazar P'arpec'i and Elišē. However, a few graffiti and inscriptions and a papyrus containing a sort of Greek phrasebook written in Armenian script are the only tangible monuments from the fifth century (see [Orengo 2017](#): 1031–4). The literary sources are only transmitted in much later manuscripts, the oldest of which go back to the late ninth century, which means that we cannot really be certain that they faithfully reflect the actual language spoken at least 400 years earlier.

Besides the classical learned and religious language that was still in use, a new written standard, based on western dialects, was created to serve the practical purposes of the state of Cilicia during the thirteenth and fourteenth century, but after the fall of the Armenian kingdom in 1375, there was no administrative system to support a written norm adapted to the spoken language. From the seventeenth century, a lingua franca, *vačarakanakan hayerēn* ‘merchant's Armenian’ ([Orengo 2017](#): 1034–5), containing various dialectal features, gradually split into the two varieties of modern Eastern and Western Armenian, whose standards were fixed by the end of the nineteenth century. Of these, Eastern Armenian is the official language of the Armenian Republic, but also spoken in Arc'ax (Nagorno Karabagh) and Iran, while Western Armenian as the language of the diaspora following the genocide in 1915 survives in bilingual communities in e.g. Lebanon, Syria, Israel, France, Canada and the USA.

12.2 Evidence for the Armenian Branch

This section contains a list of phonological and morphological features that distinguish Armenian from other branches of the Indo-European family.

12.2.1 Phonological Innovations

The most important phonological innovations characterizing the Armenian branch are listed below.¹

Vowels and Semivowels

1. Raising of long **ē* and **ō* to *i* and *u* (written *ow*) respectively, cf. *sirt* ‘heart’ < **kērd-*, *towr* ‘gift’ < **doh₃ro-*.
2. Raising of short **e* and **o* to *i* and *u* before nasals, cf. *hin* ‘old’ < **seno-*, *cown-r* ‘knee’ < **gonu-*.
3. Loss of basic length opposition for all vowels: **ā*, **ī* and **ū* merge with their short counterparts, cf. *mayr* ‘mother’ < **mah₂tēr* and *acem* ‘lead, bring’ < **h₂aġ-e-*.
4. Merger of front diphthongs **ei/*oi* into *ē* (a mid-high, eventually short vowel, distinguishing from the more open *e*), cf. *e-dēz* ‘piled up’ < **(h₁)e-d^heiġ^het*, *mēg* ‘cloud’ < **h₃moig^ho-*. While **ou* yields *oy*, cf. *boys* ‘plant, herb’ < **b^hou(h₂)ko-*, the usually assumed parallel merger of back diphthongs **eu/ou* > *oy* may not be correct. Thus, Lamberterie (1982: 81–82) assumes a development **eu* > *iw*, e.g. *hiwcanim* ‘pine away’ < **seuġ-/seuġ-* (OE *sēoc*, Goth. *siuks*). See also Olsen 2020.
5. Loss of tonal accent and fixation of stress, at first on the penultimate syllable, eventually leading to syncope of all final syllables. With few exceptions, stress is thus synchronically fixed on the final syllable.
6. At a later stage than (5), weakening of unstressed high vowels and diphthongs, whereby *i* and *u* become [ə] (usually unwritten), *ē* becomes *i*, *oy* becomes *u*, while *ea* becomes *e*.² Compare e.g. nom.sg. *sirt* ‘heart’, gen. *srti* [sər’ti]; *sēr* ‘love’, gen. *sirōy*; *loys* ‘light’, gen. *lowsoy*; *arāk’eal* ‘messenger, apostle’, gen. *arāk’eloy*.
7. Vocalic resonants **r*, **l*, **m*, **n* generally yield *ar*, *al*, *am*, *an*, cf. *mard* ‘man, mortal’ < **mr̥tó-*, Gr. (Aeol.) *βροτός*, cf. also Ved. *mṛtá-* ‘dead’.
8. While intervocalic **i* is lost, like in e.g. Greek, the reflex in initial position is not clear. Options include:
 - a. *ǰ-* as in *ǰowr* ‘water’ < **iuHr-o-*, Lith. *jūra* ‘sea’

¹ For various attempts at establishing a relative chronology of the Armenian sound changes, see Kortlandt 1980a; Ravnæs 1991; Job 1995. A recent summary of Armenian historical phonology is presented by Macak (2017). See also the general surveys by Meillet (1936); Solta (1963); Godel (1975); Schmitt (1981); Lamberterie (1989); Olsen (2017b).

² The diphthong *ea* results from both **ea* and **ia* arising after the loss of intervocalic consonants.

- b. *j-* as in *jow* ‘egg’ < **iōjio-* vel sim
 c. zero as in *nēr* ‘daughter-in-law’, Lat. *ianitricēs*.³ Perhaps also *ors* ‘hunt, game’ if < **iorko-* (thus Martirosyan 2010: 706).

An apparent reflex *l* should probably be explained by other processes. In *leard* ‘liver’ < **iek^wɫt*, contamination with **leip-* ‘fat, lard’ is conceivable, cf. OHG *lebara* ‘liver’. Similarly, the word *lowc* ‘yoke’ could have been secondarily affected by the verb *lowcanem* ‘to loosen, untie’.

9. Initial **u-* yields *g-*, cf. *get* ‘river’ < **uēd-os-*. The internal outcome is more complex and alternates between *g*, *w* and zero.⁴ It is possible that these reflexes result from a relatively late phonemic split of an intermediary **y^w*, which seems to be indirectly attested in Georgian *yvino* ‘wine’, if borrowed from an earlier form of Arm. *gini* ‘id.’ < **uojn-io-*. Note also Geo. *yvia* ‘juniper’, Arm. *gi* ‘id.’ (HAB 1: 554).

Laryngeals

10. Loss of consonantal laryngeals would be consistent with the development in the other non-Anatolian languages and thus not a specific Armenian feature. It has been claimed that initial **h₂-* and **h₃-* are preserved as *h-* before an original *e*, e.g. *haw* ‘bird’ < **h₂eui-*.⁵ There are, however, a number of problematic counterexamples, and the hypothesis requires several ad hoc reconstructions (Olsen 1999: 766–7; Clackson 2005: 155; Macak 2017: 1059).
11. Laryngeal vocalization in initial position (“prothetic vowel”) before consonants except **u*, cf. *astl* ‘star’ < **h₂stēl* for **h₂stēr*. It is debated whether Armenian, like Greek, shows a triple representation, but the evidence for this claim, most prominently *inn* ‘nine’ if < **h₁neun*, is scarce.⁶ Besides, triple representation of the prothetic vowels would be at variance with the development in other positions.
12. Vocalization of all laryngeals to *a* between consonants in initial and final syllables, cf. *keraw* (aor.act.3sg.) ‘ate’ < **g^werh₃-to*. In internal syllables the conditioning of vocalization versus loss is not fully clear (Olsen 1999: 767–8).
13. Double vocalization of **RHC* > *aRaC*, cf. *haraw* ‘south’ < **prh₃uV-*.
14. Vocalization of at least **h₂* after **i/u* in auslaut as in Greek, cf. *sterj* ‘sterile’ < **steria-* < **ster-ih₂-*. It cannot be excluded that this was a morphologically

³ The exact reconstruction is difficult, but perhaps **(h)jenh₂tēr* > **(h)jentēr* (deletion of internal laryngeal) > **(h)jinēr* (*-en- > *-in-; *-nt- > -n-) > *nir-* (**ē* > -i-; syncope of unaccented *-i-) → analogical nom.sg. *nēr*, cf. the pattern *sēr*, *siroy* ‘love’ (Olsen 1999: 190–1).

⁴ For a discussion of the conditioning, see Eichner 1978: 148–9; Olsen 1986; Ravnæs 1991: 72–3; Matzinger 1992; Olsen 1999: 787–8.

⁵ Thus Austin (1942: 22–3), followed by Winter (1965), Greppin (1973), Kortlandt (1980b), Martirosyan (2010: 712–13) and others.

⁶ Triple representation is advocated by e.g. Winter (1965), Kortlandt (1987), Beekes (1988, 2003), and Martirosyan (2010: 765–6). The opinion that all vocalic laryngeals yield *a* is defended by Klingenschmitt (1970: 80 and 1982: 105), Olsen (1985 and 1999: 262–4), Lindeman (1987: 75–83), and others.

motivated change, i.e. a levelling in favour of the oblique cases where $*\text{-}i\text{-} < *\text{-}i\text{-}ah_2\text{-}$. On the other hand, there is evidence to suggest vocalization of internal $*\text{-}ih_2/3\text{-}$ and $*\text{-}uh_2/3\text{-} > *\text{-}i\text{-}/*\text{-}u\text{-}$ as well (cf. Olsen 1992; 1999: 770–1), similar to the “breaking” in Greek and Tocharian (cf. Section 12.4.1), though this is not widely accepted.

Other Consonants and Clusters

15. Primary palatalization: the PIE palatals $*\hat{k}$, $*\hat{g}$ and $*\hat{g}^h$ yield *s*, *c* and *j* respectively.
 - a. At an earlier stage, (labio)velars had become palatals after **u* (including *u*-diphthongs), cf. *dowstr* ‘daughter’ $< *d^hugh_2t\bar{e}r$, *loys* ‘light’ $< *le/o_1ko\text{-}$.
16. Chain shift of the remaining PIE stops:
 - a. PIE voiceless stops $*t$ and $*k$ become t' and k' respectively, while $*p$ usually becomes *h* (via $*p^h$ and/or $*f$), disappearing before *o*, cf. *het* ‘footstep’ $< *pedom$ vs. *otn* ‘foot’ $< *podm$.
 - b. PIE voiced stops $*b$, $*d$ and $*g^{(w)}$ become *p*, *t* and *k*.
 - c. PIE voiced aspirated stops $*b^h$, $*d^h$ and $*g^{(w)h}$ become *b*, *d* and *g*.
17. Lenition or loss of particular voiceless and voiced aspirated stops. The circumstances are complex, but at least the following developments are fairly certain:
 - a. intervocalic $*p$ and $*b^h > w$, cf. *ew* ‘and’ $< *h_1epi$, $-(a)wor$ ‘carrying’ $< *b^h\text{-}orah_2\text{-}$
 - b. intervocalic $*t > y$ before front vowels, cf. *hayr* ‘father’ $< *ph_2t\bar{e}r$; intervocalic $*t > w$ before back vowels, cf. *cnaw* (aor.3sg.) ‘was born’ $< *(e\text{-})\hat{g}enh_1\text{-}to$; when not following the stressed syllable, intervocalic $*t$ disappears entirely, cf. *č'ork* ‘four’ $< *k^{w\text{et}}\acute{o}res$
 - c. intervocalic $*\hat{g}^h > z$, cf. *lezow* ‘tongue’ $< *lei\hat{g}^h\text{-}uh_2\text{-}$
 - d. intervocalic $*g^{wh} (> *j) > \acute{z}$ before front vowels, cf. *iž* ‘snake’ $< *h_1\acute{e}g^{wh}\text{-}i\text{-}$ (apparently no examples of $*\text{-}g^h\text{-}$)
 - e. internal $*\text{-}pt\text{-} > \text{-}wt'\text{-}$, cf. *ewt'n* ‘seven’ $< *septm$
 - f. internal $*tR$, $*kR$, $*\hat{k}R > wR$, cf. *arawr* ‘plough’ $< *h_2arh_3tro\text{-}$, *mawruk* ‘beard’ $< *(s)mo\hat{k}ru\text{-}$
 - g. internal $*\text{-}pn\text{-} > \text{-}wn\text{-}$, cf. *k'own* ‘sleep’ $< *suopno\text{-}$
 - h. initial voiceless stops are lost before resonants, cf. *li* ‘full’ $< *pleh_1to\text{-}$
 - i. initial $*pt\text{-} > t'\text{-}$, cf. *t'er* ‘side; leaf’ $< *pter\text{-}$.⁷
18. Secondary palatalization of (labio)velars. This development is most clearly seen in *č'ork* ‘four’ $< *k^{w\text{et}}(u)ores$ and *jerm* ‘warm’ $< *g^{wh}ermo\text{-}$.⁸ This

⁷ The seemingly missing lenition of $*k^{(w)}$ and $*g^{(w)h}$ (cf. Kortlandt 1980a; Kümmel 2017) and the outcome of lenited $*d^h$ (*z* or *r*, cf. Jasanoff 1979: 143–4; Martzloff 2016) are subject to debate.

⁸ There are no examples involving $*k$, $*g^h$ or $*g^w$. Considering the evidence at face value thus leaves an asymmetrical pattern, which is why it is sometimes assumed that palatalization affected all velars (Kortlandt 1975). Numerous exceptions such as *keam* ‘to live’ $< *g^{w\text{ie}}h_3\text{-}$ would thus require analogical explanations which are not always straightforward.

- feature is perhaps not exclusively Armenian (cf. Section 12.4.3), but another uniquely Armenian rule, the “*awcanem*-rule” (Kim 2018: 258) proves the preservation of labiovelars into the immediate prestige of Armenian: $*VnK^w > *Vw\hat{K}$ (cf. 15. a), e.g. $*h_3ng^w- > awc(anem)$ ‘anoint’.
19. While the general reflex of $*s$ is $h/\emptyset$ much like Greek, conditioned developments are subject to more controversy.
 - a. To explain the usual nominal and pronominal ending of the nom.pl. $-k'$, it is suggested by e.g. Pedersen (1905: 209–227) and Kortlandt (1984) that it is the regular outcome of final $*-s$.
 - b. A *ruki*-like development of final $*-s > -r$ after i and u (including $*\bar{e}$ and $*\bar{o}$ following [1]) may explain intricacies such as singular aorist imperatives like *towr* ‘give’, which could then reflect the original injunctive $*doh_3-s$ (cf. Pedersen 1905: 228; Olsen 1989).
 20. Metathesis in clusters of voiced (aspirated) stops and resonants whereby e.g. $*-dr-$, combined with the sound shift (16), yields $-rt-$ with initial vowel prothesis, cf. *artawsr* ‘tear’ $< *dra\hat{ku}-$, *merj* ‘near’ $< *me-\hat{g}^h sr-i$.
 21. Epenthesis of $*i$ and $*u$ caused by an $*i$ or $*u$ in the following syllable, cf. *ayl* ‘other’ $< *h_2 ali\bar{i}o-$, *awl-i* ‘strong alcoholic drink’ $< *h_2 alu-$. While these changes are not spontaneous, the conditions are not fully clear. It seems that i -epenthesis only took place before resonants and after the vowels a and o while u -epenthesis was restricted to a rather different environment, also after i (perhaps e) and before stops, cf. *giwt* ‘discovery’ $< *\bar{u} id-(t)u-$. On the other hand, it is not found in well-established u -stems such as *asr* ‘wool’ $< *p\hat{ku}-$ and e.g. Beekes (2003: 205) is sceptical of its existence altogether. Perhaps the original place of accent played a role in the development of u -epenthesis (see Olsen 1999: 798–801 with references).
 22. Particular developments of various clusters including
 - a. $*sK$, $*Ks > c'$ in most cases, cf. *c'elowm* ‘split, break’ $< *skelH-$; *vec* ‘six’ $< *suue\hat{ks}$. Initially, the outcome $\check{s}$ - may sometimes be observed, and might be the result of palatalization before front vowels. Alternatively, Martirosyan (2010: 516) suggests that $\check{s}$ - regularly develops from $*sKHV-$ as opposed to $*sKV- > c'-$. It is debated whether $-\check{c}'-$ is the palatalized version of $*-sK-$ in internal position or should be derived from $*-sK\bar{i}-$.
 - b. $*d^hi > \check{j}$, cf. *měj* ‘middle’ $< *med^hi o-$. The outcome of $*t\bar{i}$ and $*d\bar{i}$, either c'/c or $\check{c}'/\check{c}$, is more controversial (see e.g. Olsen 1993, Kocharov 2019: 30–1).
 - c. $*R\bar{i} > R\check{j}$, cf. *sterj* ‘sterile’ $< *sterih_2-$.
 - d. $*s\bar{u}$, $*t\bar{u} > k'$, cf. *k'oyr* ‘sister’ $< *sues\bar{o}r$.
 - e. $*d\bar{u} > (V)rk-$, cf. *erkow* ‘two’ $< *du\bar{o}$.⁹

⁹ Others favour a regular development $*d\bar{u} > k$, cf. Beekes 2003: 199–200. For a more exhaustive overview of developments in clusters, see Godel 1975: 78–9.

12.2.2 Morphological Innovations: The Verb

The Armenian verb has undergone a number of morphological simplifications, such as loss of the dual and the distinction between an optative and a subjunctive, while the perfect only survives in synchronically opaque relics.¹⁰ Specific Armenian changes include

23. Generalization of *-e-* as thematic vowel with the exception of the subj. 1pl. *-owk' < *-omes* and the participle in *-own < *-ont-/*-omh₁no-*.
24. Merger of the thematic (or *e*-stem) endings and the verb 'to be' in the present active, thus *berem* 'I carry' like *em* 'I am'.
25. Creation of a mediopassive paradigm in *-i-* from statives in **-eh₁-*.
26. Creation of a new imperfect preterite.
27. Merger of old aorist and imperfective stems for the formation of "root aorists".
28. Creation of a "weak" aorist stem in *-c'-*, possibly a remodelling of the old *s*-aorist (cf. Klingenschmitt 1982: 286–7; Olsen 2017b: 443).
29. Formation of a subjunctive morpheme *-ic'-* of disputed origin.
30. Formation of a causative in *-owc'anem*, aor. *-owc'i*, also of disputed origin.
31. Formation of a voice-indifferent infinitive in *-l < *-lo-*.
32. Formation of a past participle in *-eal* (*o*-st.), similar to the Slavic *l*-participle.

12.2.3 Morphological Innovations: The Noun

In the noun, the categories of grammatical gender and the dual number are lost, while an inventory of seven cases is maintained despite several cases of syncretism. The most notable inflectional innovations include

33. Formation of a gen.dat.abl. plural in *-c'*, e.g. *i-st. srtic'* from *sirt* 'heart', possibly originally an adjective in **(i)-s₁kō-*.
34. Introduction of a new abl.sg. ending *-ē*, probably *< *-eti*.
35. Introduction of a new loc.sg. ending *-i* (*a-*, *i-* and sometimes *o*-stems), probably *< *-h₁en*.
36. Merger of old root nouns, heteroclitics and *s*-stems with other stem classes.
37. Creation of a heteroclitic *u-/n*-stem paradigm from original *u*-stem adjectives, e.g. *barjr* 'high', gen. *barjow*, nom.pl. *barjownk'*: Hitt. *parku-*.
38. Creation of a marginal *l*-stem paradigm, apparently extended from the paradigm for 'star', *astl*.

From the field of nominal word formation, the most remarkable innovation must be:

39. The creation of a complex abstract noun suffix *-owt'iwn* on the basis of inherited elements.

¹⁰ For more elaborate treatments of morphological innovations, see Klein 2007; Olsen 2017a; 2017b; Klingenschmitt 1982 on the verb; Olsen 1999 on the noun; Matzinger 2005a on nominal inflection.

12.2.4 Morphological Innovations: The Pronoun

The pronoun is notoriously a word class that is subject to changes and analogical remodellings, and here Armenian is no exception. However, one feature is particularly characteristic:

40. A systematic distinction between three deictic markers: *s* for the first person, *d* for the second and *n* for the third. This system includes the postponed articles, *-s*, *-d*, *-n*, the anaphoric pronoun *sa*, *da*, *na*, the demonstrative *ays*, *ayd*, *ayn* and various other pronouns, adverbs and interjections.

12.2.5 The Lexicon and Remaining Innovations

The most remarkable feature of the Armenian lexicon is the scarcity of inherited lexemes seen in relation to the abundance of loanwords, mostly from Middle Iranian sources, and words of obscure origin. The etymological background of around 50 per cent of the Armenian vocabulary is unknown, and thus an abundance of words that are only attested in this branch help to define Armenian as an independent member of the Indo-European family.¹¹

12.3 The Internal Structure of Armenian

Armenian is generally considered to be a single-language branch and indeed, Classical Armenian appears to be a highly standardized language with very few traces of the dialectal diversity that is likely to have existed at the time of the composition. According to Meillet (1904), the later dialects all derive from a uniform learned *κοινή* with very few modifications. As examples of dialectal archaisms, Meillet himself (also 1936: 11) mentions the original dialectal form *lizow* ‘tongue’ vs. Classical *lezow* with umlaut *i-u > e-u* and the preservation of the accusative marker *z-*, mostly lost in the later language, but preserved in the dialects around Lake Van. Within the Classical language itself, we also find doublets such as *t’aršam/t’arām* ‘withered’. Another indication of early dialectal differentiation is the word *ays*, usually ‘evil spirit’, but also attested in the primary meaning ‘wind’ in Eznik, who explicitly calls it a word of the southerners (Clackson 2005: 154). The fifty to sixty modern Armenian dialects all fall into one of the two main groups, Western and Eastern, with further subgrouping possible. Some important criteria for the classification of dialects are the reflection of the Classical Armenian stops and the formation of the present indicative where both Western and Eastern Armenian employ innovative but different formations.¹²

¹¹ See the excellent overview by Clackson 2017.

¹² On the topic of dialectal subdivision and the question of dialectal diversity in the earliest literature, see Adjarian 1909; Martirosyan 2010: 689–704; Martirosyan 2018; Weitenberg 2017.

12.4 The Relationship of Armenian to the Other Branches

In the pre-literary period, there must have been close linguistic contact between Armenian and a great number of other known and unknown languages, Indo-European – especially shown by the massive layer of Middle Iranian loanwords – as well as non-Indo-European, of which the non-Indo-European element is responsible for a substantial part of the lexicon, cf. e.g. *xnjor* ‘apple’ : Hurrian *hinzuri* ‘id.’. While there are relatively few borrowings from Kartvelian in the oldest language, the areal influence of the Kartvelian languages may explain the dialectal glottalization of old mediae.¹³ On the syntactic level, the ergative-like construction with participles in *-eal* where the agent is in the genitive and the direct object in the accusative, e.g. *nora* (gen.) *gorceal ē z-gorc* (acc.) ‘he has done the work’, likewise finds parallels in Kartvelian (Stempel 1983: 80–7), but also in Iranian, however (Meyer 2017: 109–60).

Occasionally, it seems justified to attribute lexemes exhibiting irregular sound change to an unidentified Indo-European language. Thus *bowrġn* ‘tower, pyramid’ and *dowrġn* ‘potter’s wheel’ have the appearance of derivatives of **b^herġh-* ‘(be) high’ and **d^herġh-* ‘run’ respectively, but in both cases the root vocalism and the *centum* reflex of **-ġh-* are at variance with established Armenian sound laws.

Otherwise, Armenian shows the strongest similarities to the group of Balkan languages, Phrygian, Albanian and in particular Greek (see Figure 12.1). Some interesting features of this group are shared with Indo-Iranian (in particular the augment and the prohibitive adverb **meh₁*) and a few with Tocharian.

12.4.1 Armenian and Greek

The idea of a particularly close relationship between Armenian and Greek has a long history. Thus Pedersen (1905; 1924) mentioned a number of Greek-Armenian isoglosses and concluded that no other language was as close to Armenian as Greek. Later Bonfante (1937) provided a long list of phonological

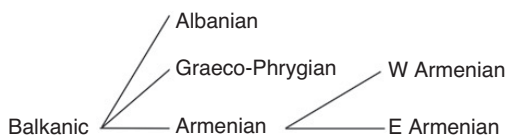


Figure 12.1 The position of Armenian

¹³ Adherents of the “Glottalic Theory” interpret this characteristic feature as an archaism (e.g. Gamkrelidze 2003 with references).

correspondences, most of them not exclusively Graeco-Armenian, Hamp (1976) referred to the “growing list of Greek-Armenian isoglosses”, concluding that the time was “approaching when we should speak of Helleno-Armenian”, and Lamberterie (1983) considered Armenian to be particularly close to Greek.

The opposite stand was taken by Clackson (1994: 199–200), who ended his investigation with the following negative conclusion: “The absence of any compelling explanation of a morphological development of either language suggests strongly that the languages did not form a sub-group.” Even the impressive number of lexical correspondences was toned down: allegedly, only five word-pairs might reflect a common agreement made jointly by Greek and Armenian.

Most recently, Kim (2018) discarded most of the lexical correspondences as “general root cognations, not full word equations” and the notion of a Graeco-Armenian unity as an example of the “inertia of established scholarly opinion”.

However, while the lexical correspondences are certainly the most prominent, generally dismissing phonological and especially morphological correspondences seems unwarranted. In fact, a number of early phonological innovations in Armenian appear to be shared with Greek.

This goes for certain patterns of laryngeal vocalizations, particularly in initial position before consonant (11), in connection with the vowels **i* and **u* (14) and of “long resonants”, i.e. **CRHC* clusters. As for the initial vocalization, Greek clearly shows a triple reflex (*ε/α/ο*) of vocalized laryngeals, while this outcome is far from assured for Armenian. In fact, one typically finds *a* in place of both **h₂* and **h₃*, thus *astl* ‘star’ = Gr. ἀστήρ; *aniw* ‘wheel’ ≈ Gr. ὀμφαλός ‘navel’. Indisputable examples involving **h₁* are unfortunately lacking (see e.g. Clackson 1994: 35).¹⁴ At any rate, the tendency for initial laryngeal vocalization is not found anywhere else, apart from Phrygian (Section 12.4.2), and it may to some extent be regarded as a shared innovation.

A closely related change concerns the Greek development of **Cih_{2/3}C* > **Cīā/ōC* and **Cuh_{2/3}C* > **Cūā/ōC*, which operated in originally unaccented syllables, as observed in e.g. Gr. ζῶός ‘alive’ < **gʷīōwó-* < **gʷīh₃-uó-*.¹⁵ In

¹⁴ However, Clackson (1994: 35) considers a single reflex *a-* most likely on theoretical grounds. The final decision depends on the exact analysis of *atamn* ‘tooth’, traditionally derived from the root **h₁ed-* ‘eat; bite’ (or ‘gnaw’?) and *anown* ‘name’.

¹⁵ See Francis 1970: 276–7; Normier 1977: 182 n. 26; Rasmussen 1991; Clackson 1994: 41–9; Hyllested 2004; Olsen 2009 (for the conditioning); Woodhouse 2015. While this rule, sometimes referred to as “laryngeal breaking” or “Francis’ Law”, has not met with universal acceptance, it remains, in our view, the most economical solution to a number of etymological issues. The only serious counterexample, viz. Gr. θῆμός ‘spirit’ (cf. Chapter 11), may be illusory. As suggested by Kristoffersen (2019), the Greek word, like OHG *tuom* ‘vapour’ and Lat. *fūmus* ‘smoke’ (without Dybo’s Shortening! Cf. Section 9.2.3), seems to represent an

Armenian, the operation of a similar rule, $*-ih_{2/3-} > *-i\bar{d}- > *-i\bar{a}-/*-uh_{2/3-} > *-u\bar{d}- > *-u\bar{a}-$, is suggested especially by *erkar* ‘long’, which is identical to Gr. *δηρός* ‘id.’ $< *duh_2-r\acute{o}-$. The value of this example has been questioned due to the possible contamination of the adverb $*duah_2m$ ‘far’ (Hitt. *tūān* ‘to this side’, *tūuaz* ‘from afar’ and Gr. *δίην* beside the morphologically aberrant Arm. *erkayn*), but there is in fact more Armenian material to suggest that this rule was regular (see Olsen 1992; 1999: 770–3). Note e.g. *keam* ‘to live’ $< *g^{wih}_3u-$, which is traditionally difficult to reconstruct (see Martirosyan 2010: 356–7). The development of these $*CI/UHC$ sequences may be somehow connected with the rather complex and poorly understood development of $*CRHC$ clusters in both Armenian and Greek (Woodhouse 2015). However, as laryngeal breaking is a well-established feature of Tocharian, it can hardly be considered an exclusive Graeco-Armenian isogloss.

It has been suggested (Olsen 1989) that Greek and Armenian share a tendency to voice posttonic $*Nt > Nd$, though the contexts are not identical as the development in Greek is restricted to $*Nt$, e.g. *δέκα*, *δέκατος* ‘ten’ vs. *δεκάς*, *δεκάδος* ‘a decade’, but $*h_1énterah_2-$ ‘entrails’ $>$ Arm. *ənderk*’ vs. Gr. *έντεπα*. Rather than an actual shared innovation, we may be dealing with an areal feature.

In general, the most significant argument in favour of a common intermediate proto-language is the existence of shared morphological innovations. For Greek and Armenian, at least a handful of cases of this kind may be adduced:

- formation of a *nu*-present $*ues-nu-$ from the root $*ues-$ ‘dress’: Arm. *z-genowm*, Gr. *έννυμι* as a common substitution for the causative $*uos-éie-$ (Klingenschmitt 1982: 248)
- formation of a reduplicated aorist $*ar-ar-e/o-$: Arm. *arari* ‘I made’, Gr. *ήραρον* ‘I fixed’ (Chapter 11)
- formation of a (reduplicated?) present stem $*(si)-slh_2-s\hat{k}e-$: Arm. *alač’em* ‘ask, request’, Gr. *ίλάσκομαι* ‘appease’ (Klingenschmitt 1970). The development $*-s\hat{k}- > -č’-$ seems to be regular before front vowels, and the reduplicative syllable would be lost due to syncope in Armenian. While the root is not exclusively Graeco-Armenian (cf. e.g. Lat. *sōlor* ‘console’), the stem formation, perhaps patterned on $*\hat{g}i-\hat{g}h_3-s\hat{k}e-$ (Arm. *čanač’em*, Gr. *γιννώσκω*), is unique for the two branches
- inflection of the $*-men(t)$ -stems: Arm. *sermn*, gen. *serman*, Gr. *σπέρμα*, *-ματος* ‘seed’, Arm. *žermn*, gen. *žerman* ‘heat, fever’. Greek and Armenian seem to have shared the generalization of the suffix variant $*-m\hat{n}t-$ in this type, which is thus a likely candidate for a common innovation¹⁶

o-grade, $*d^{o}ou(h_2)mo-$ (Gr. $*-V\bar{u}- > -\bar{u}-$ before labials) as opposed to the zero grade of Ved. *dhūmá-*, Lith. *dūmai*.

¹⁶ Unstressed $*-m\hat{n}t-$ $> -man-$. However, an analogical explanation of the Armenian paradigm cannot be definitely excluded.

- creation of the grammaticalized adjectival suffix conglomerate **-ōdēs* < **-o-h₃od-ēs*, lit. ‘smelling’, e.g. Arm. *awazowt* : Gr. ἀμαθώδης ‘sandy’
- formation of the suffix conglomerate **-e(h₁)u- + -to/ah₂-* or *-ti-* in Arm. *-oyt’* < **-e(h₁)u-ti-*, e.g. *erewoyt’* ‘appearance’, Gr. τέλευτή ‘end’ < **-e(h₁)u-tah₂-*. The Greek type in *-ενσις* is late, but a common prestige is most likely a shared innovation.

The most spectacular evidence for a Graeco-Armenian subgroup remains a set of lexical isoglosses which vary in nature. Some are simple exclusive root correspondences, but the following etyma are among the strongest examples showing common morphological and/or semantic innovations based on inherited roots. For a comprehensive collection of material, see e.g. [Solta 1960](#), [Clackson 1994](#) and [Martirosyan 2013](#).

- **mēdesa-* ‘mind’: Arm. *mit*, usually pl. *mit-k’* (gen.-dat.-abl.pl. *mt-ac’*); Gr. *μῆδεα* ‘counsels, plans, arts’, cf. *μῆδομαι* ‘to contrive, plan’. At least the long root vowel, whatever its explanation, seems to be an innovation.¹⁷ Note also the similar semantics as opposed to Umb. *meřs* ‘law’. The long root vowel cannot be the reflection of an original Narten-ablaut (pace [Clackson 1994](#): 148) since Gr. *μῆδομαι* only has middle forms. Also, the long vowel forms found in Germanic and Old Irish are most likely secondary ([Meissner 2006](#): 80–1).
- **d^heh₁s-* ‘god’: Gr. *θεός* ‘god’ (< **d^heh₁s-o-*) agrees semantically with Arm. *di-k’* ‘(heathen) gods’ (< **d^heh₁s-es*) as opposed to Lat. *fēriae* ‘holidays’, *fānum* ‘temple’ which, together with potential Anatolian cognates, viz. HLuw. *tasan(-za)* ‘votive stele’, Lyc. *ṡṡēn-* ‘altar’, suggest an original meaning ‘votive, sacred (thing)’. This would make the semantic change to ‘god’ a shared innovation ([Lamberterie 2013](#): 35–6) in which Phrygian also takes part, cf. Phryg. (dat.pl.) *δεως* ‘god’ ([Section 12.4.2](#)).
- **mr̥tō-* ‘mortal’: Arm. *mard* ‘(mortal) man, person’, Gr. (Aeol.) *βροτός* ‘mortal’. Formally, this is obviously the past participle of PIE **mer-* ‘to disappear, to die’. The semantic shift from ‘dead’ (Skt. *mṛtá-*) to ‘mortal’, presumably a contrast formation to the privative **n-mr̥to-* ‘immortal’, is not a very trivial innovation and has a low chance of reflecting parallel developments. It is also remarkable that the contrast human : god is expressed by the same word pair, Arm. *mard* : *dik’*, Gr. *βροτός* : *θεός*.
- **suekura-* ‘mother-in-law’: Arm. *skesowr*, Gr. *ἐκυρά*. Presumably this exclusive Armenian-Greek form replaced the more archaic feminine **suekruh₂-* (cf. Skt. *śvaśrū-*, Lat. *socrus*, OCS *svekry*) by analogy with **suekuro-* ‘father-in-law’ (itself probably a secondary derivative of PIE age, see [Olsen 2019](#): 153). Although this innovation may be said to be trivial, it is not found elsewhere, where the original *uh₂-*stem is generally well preserved.

¹⁷ It may result from contamination with **meh₁-* ‘measure’ (GEW 2: 223).

- **mātru(u)jah₂-* ‘stepmother’: Arm. *mawrow*, Gr. *μητρονία*. Armenian and Greek agree in derivation and meaning as opposed to OE *mōdrige* ‘mother’s sister’. It is uncertain whether the Germanic forms reflect the same derivation. Clackson (1994: 145–7) considers this isogloss insignificant since both the form and meaning might be archaic (see also Olsen 2019: 156–7). On the other hand, the agreement of an exclusive form and meaning ‘stepmother’ as opposed to the expected ‘mother’s sister’ in Germanic is striking enough to suggest a joint innovation.
- **prei̯s-g^wh₂-u-* ‘one who goes in advance, elder’: Arm. *erēc*, gen.sg. *erīc’ow*; Gr. *πρέσβυς*, Cretan *πρεῖσγυς* (Lamberterie 1990: 909–11, Clackson 1994: 165; on the phonology, see Olsen 1988). Lat. *prīscus* ‘ancient’, an *o*-stem, is unlikely to continue an older *u*-stem and rather reflects the suffix **-ko-*, cf. Weiss 2020: 315.
- **osara-* ‘harvest’: Arm. (*amis*) *ara-c* ‘the sixth month of the ancient Armenian calendar (month of harvest)’ and Gr. *ὀπ-ώρᾱ* ‘part of the year between the rising of Sirius and of Arcturus, between summer and autumn’. The shared preform **osara-* (or **ohara-* if **s > h* was a shared development) seems to be a thematization of the PIE strong stem **h₁os-r-*, cf. Ru. *ósen* ‘autumn’, Goth. *asans* ‘harvest’ (Martirosyan 2013: 110).
- **g^wh₂(a)no-* ‘acorn’: Arm. *kalin*, Gr. *βάλανος* (Clackson 1994: 135). Greek and Armenian are the only branches to agree on the suffix, cf. Lat. *glāns* (< **g^wh₂-nd^h-*), RuCS *želudb* (< **g^welh₂-ond^h-*), Lith. *gilė* (< **g^wh₂-jah₂-*).
- **perHi-men-* ‘piercing object’: Arm. *heriwn* ‘awl’ < **perHimōn*, Gr. *περόνη* ‘pin, buckle, brooch’ < **perHimneh₂*, cf. *ἀκόνη* ‘whetstone’: *ἄκμων* ‘anvil’. It may be assumed that the root is **perH_i-*, which would explain Gr. *πείρω*, OCS *na-perǫ* ‘pierce’ as simple thematic presents (Olsen 1999: 492). Of course, it cannot be excluded that this isogloss is a shared archaism.
- **pseud-* ‘lie’: Arm. *sowt* ‘false’, stem ‘lie’, Gr. *ψεύδομαι* ‘deceive, lie’, *ψεῦδος* ‘lie’ (Clackson 1994: 168–9). If the basic root is **pseu-* ‘blow’, as suspected by Taillardat (1977: 352–3; cf. Fr. *vendre du vent*, Eng. *windy*, hot *air*), only Armenian and Greek agree on the root-extension *-d-* and the semantic specialization. Moreover, Arm. *sowt* < **psudo-* has the appearance of a contamination of a *ro*-adjective, like Gr. *ψυδρός*, and a full-grade *s*-stem, like Gr. *ψεῦδος*, meaning that traces of the Caland system would have survived into a common prestage. This favours a common Graeco-Armenian innovation.
- **megh₂r-* ‘make great’: Arm. *mecarem* ‘honour’, Gr. *μεγαίρω* ‘grudge, envy’. The denominative verb based on the *r*-stem variant of the heteroclititic corresponding to Ir. **mazar-/mazan-* or **masar-/masan-* (Kümmel 2012) is almost certainly a common innovation.
- **drep-ñnah₂-* ‘sickle’ in Gr. *δρεπάνη* ‘sickle’, Arm. *artewan* (-ownk’, -anc’/ -ac’) ‘eyelid; brow’ (Lamberterie 1983: 21–2). The root **drep-* is not

exclusively Graeco-Armenian, thus Ru. *drápat* ‘scratch, tear’ beside Gr. *δρέπω* ‘pluck, cut off’, but the striking correspondence consists in the derivational chain **drep-mḡ* (Gr. (Hsch.) *δρέμμα· κλέμμα* (about stealing fruit); Beekes 2010: 353) ⇒ **drep-ḡnah₂-* > *artewan-/δρεπάνη*, very much in accordance with inherited principles. Clackson’s tentative suggestion (1994: 112) of a very early loan from Greek is extremely unlikely, as we have no examples of Greek loanwords borrowed before the soundshift (**d > t*).

- **h₂alh₁-trih₂-* or **h₂lh₁-trih₂-* ‘female miller’: Arm. *alawri* ‘female who grinds corn’, Gr. *ἀλετρις* ‘female slave who grinds corn’. Apparently a *vrkīh-* type derivative of an agent noun in **-ter/tor-*, an otherwise extinct derivational type in Armenian. Clackson’s suggestion (1994: 92) of “a secondary derivative of an unattested instrument noun **alawr* ‘mill’” is less economical. Again, a common innovation is the simple solution.
- **dʰal-ro-* or **dʰhl-ro-*: Arm. *dalar* ‘green, fresh’, Gr. *θαλερός* ‘blooming, fresh, abundant’. As Gr. *-λρ-* is phonotactically impossible, and Arm. *-lr-* never represents an old consonant cluster, Gr. *-επο-*, Arm. *-ar-* do not necessarily continue a sequence **-Vro-*; more likely, we are dealing with an old **-ro-* stem, only attested in Armenian and Greek. The root, however, is also found in Alb. *dal* ‘sprout, enter, come’.

Some isolated roots might be retentions from PIE but are still worth taking into account.

- **ken(-eu)-o-* ‘empty’: Arm. *sin*, Gr. *κενός*, Ion. *κεινός*, Hom. *κενεός* (cf. Clackson 1994: 138).
- **mosḡh-* ‘young bovine’: Arm. *moz-i*, Gr. *μόσχος*. Clackson’s (1994: 154) suggestion of a borrowing from Greek to Armenian seems phonetically impossible and the relatively late (eleventh century) attestation of the Armenian word is not a serious problem in itself. Most likely, it is a shared borrowing, but IE origin cannot be excluded.
- **kiuōN* ‘pillar’: Arm. *siwn*, Gr. *κίων*. The appurtenance of other cognates (cf. Lubotsky 2002; Chapter 11) is uncertain, but cannot be excluded. Clackson (1994: 140–1) considers this word a shared borrowing, which would make it an important isogloss as the forms are identical.
- The root **h₃bʰel-*, exclusively attested in Greek and Armenian, has the double meaning ‘increase’ and ‘sweep’ in both languages: Arm. *awel* ‘broom’, *awelowm* ‘increase’: Gr. *ῥοπτρον* ‘broom’, *ῥοπέλλω* ‘sweep’ (Hipponax) and ‘increase’; the verb also forms a thematic aorist in both languages: Arm. *y-awel*, Gr. *ῥοπελε* (Clackson 1994: 156–8).
- Arm. *awr* ‘day’ ~ Gr. *ἡμαρ* (cf. Chapter 11).

Finally, a number of words seem to have been borrowed at a common prestage of Armenian and Greek as the attested forms allow for reconstructions of proto-forms which, for different reasons, are unlikely to be inherited from PIE. The

shared substrate interface seems to contain several chronological layers, some presumably formed after particular Armenian or Greek sound changes.¹⁸ The following examples, where all sound changes are observed, can be considered part of the earliest layer which may have been contemporaneous with a shared Graeco-Armenian language stage.

- **aiǵ-* ‘goat’: Arm. *ayc* ‘(she-)goat’, Gr. *αἴζ*, *αἰγός*. Note the Arm. plural form *ayci-k* ‘(beside *ayc-k*)’ and derivatives *ayceay* ‘made of goatskin’, *ayceamn* ‘roebuck’ which can reflect the same **ih₂-*collective as Gr. *αἰγίς* ‘goatskin’. The etymon is probably non-IE (Solta 1960: 405; Kortlandt 1986: 38–9; and especially Kroonen 2012: 245–6). Lith. *ožỹs*, Skt. *ajā-* reflect **aǵ-* without the semivowel and although the forms are unlikely to be separated completely, the variation cannot really be explained in a PIE framework.¹⁹ In light of this, the Armenian-Greek agreement in both root structure and derivation should be considered highly significant. Another possible match is found in Alb. *edh* ‘kid’, *dhi* ‘she-goat’ < **aiǵ-iǵah₂* (Demiraj 1997: 160).
- **ant^h-r-* ‘coal, ember (?)’: Arm. *ant-el* ‘hot coal, ember’, *ant-ayr* ‘spark’ (< **ant-ari-*), dial. *ant-roc* ‘poker’; Gr. *ἄνθραξ* ‘charcoal’ (Jahowkyan 1987: 592, Martirosyan 2010: 85; 2013: 113). A substratum origin is supported by Geo. *ant-eba* ‘to burn’ and the fact that the shared root seems to contain voiceless **t^h* while there is no external support for a reconstruction **h₂antH-* vel sim.
- **sep^hs-* ‘to boil, cook’: Arm. *ep-em* ‘to cook’, Gr. *ἔπω* ‘to boil, seethe’. It is unlikely that Arm. *p* continues intervocalic **-ps-*, cf. *eres* ‘face’ < **k^wrepsah₂* (Olsen 1999: 64; alternatively Witczak 1991). Again, there are few other options than to reconstruct a voiceless aspirate, perhaps from a non-IE source.
- **tūp^h-* ‘plant, bush (?)’: Arm. *t’owp* (gen. *t’p’oy*) ‘bush, bramble’, Gr. *τύπη* ‘reed mace, *Typha angustata*’. Although the semantic details are not fully clear, and Armenian has an *o*-stem as opposed to the Greek feminine, the roots are identical. The root structure points to a substratum origin. Lat. *tūber* ‘swelling’, ON *þúfa* ‘knoll’ may be separate borrowings from the same source or entirely unrelated.
- **tarp-* ‘basket’: Arm. *t’arp* ‘fishing basket, creel’, also *t’arb* as a literary form meaning ‘wooden framework’ (HAB 2: 162; Martirosyan 2010: 281–2 with references); Gr. *τάπη* ‘large wicker basket’. There are no convincing IE etymologies (Chantraine 1999: 1095; Clackson 1994: 183; Martirosyan 2010: 281–2). This etymon may represent a very early borrowing, with the regular Armenian outcome of **tarp-* being represented in the form *t’arb*.

¹⁸ Cf. e.g. Arm. *sex* ‘melon’ ~ Gr. *σικύα* ‘bottle-gourd’ with no change of **s > h* in either language. See also Martirosyan 2013: 122–3.

¹⁹ For this reason, the connection with Av. *īzaēna* ‘leathern’ from a putative zero grade **h₂iǵ-*, mentioned e.g. by Martirosyan (2010: 58), is less likely.

Summing up, the relations between Armenian and Greek seem to be significant enough to justify a common node. They do not only consist of shallow lexical correspondences. The common morphological innovations are far from negligible, and in numerous cases, a given lexical item shows a striking similarity with respect to word formation and semantics. Exclusive loanword isoglosses further confirm this standpoint.

12.4.2 *Armenian and Phrygian*

The idea of a special relationship between Armenian and Phrygian goes back to Herodotus (7.73), who claimed that the “Armenians” (*Ἀρμένιοι*) were descendants of the Phrygians, and a quotation from Eudoxos by Stephanos of Byzantium, according to whom the Armenians come from Phrygia. He claims that their language is also very similar to that of the Phrygians. However, the closest known relative of Phrygian is undoubtedly Greek (Chapter 11), and while both Armenian and Phrygian may be attributed to the Balkan group of Indo-European of which Greek seems to be the central member, there are no exclusive isoglosses between the two.²⁰

12.4.3 *Armenian and Albanian*

Like Greek, Armenian and Phrygian, Albanian appears to belong to the Balkanic languages in the narrower sense, but apart from the palatalization of labiovelars as opposed to plain velars, perhaps a parallel development of the cluster **sʷ-* and a few lexical correspondences (Kortlandt 1986), there are hardly any conspicuous *exclusive* isoglosses between Armenian and Albanian (see further Chapter 13).²¹

12.5 The Position of Armenian

In Matzinger’s treatments of the question (2005b: 382; 2012), Greek has the central position within the Balkanic group with direct relations to Phrygian, Armenian, Albanian and perhaps – surprisingly – Tocharian.²² Evidence for the inclusion of Tocharian is extremely weak, however, and it is generally considered an entirely separate branch of Indo-European (see Chapter 6). Evidence for the Balkanic group is found at all levels, phonology, morphology and lexicon, and can be summarized as follows:

- “laryngeal breaking” (14): Greek, Armenian and Tocharian

²⁰ See Matzinger 2005b and 2012 for details.

²¹ Details on the connection between Armenian and Albanian are presented by Kortlandt (1986).

²² See e.g. also Klingenschmitt 1994 and the somewhat idiosyncratic overview by Holst 2009.

- development of at least $*-ih_2 > *-i\partial_2$ (14): Greek, Armenian and Albanian (Klingenschmitt 1994: 244–5)
- prothetic vowels (11): Greek, Phrygian and Armenian; Greek and Phrygian agree on “triple representation”
- traces of labiovelars in *satem* languages. In Armenian and Albanian, old voiceless and voiced aspirated labiovelars seem to palatalize (Pisani 1978), and a similar tendency may be observed in the *centum* language Greek, where labiovelar mediae typically avoid palatalization, cf. e.g. Arm. *keam* ‘live’ : Gr. *βέομαι*, *βίωτος*. Here we seem to be dealing with an areal feature
- loc.pl. ending $*-si$ for $*-su$: Greek, Albanian; the origin of Arm. $-s$ is unknown
- mid.1sg. primary ending $*-mai$ for original $*-h_2a\check{i}$: Greek ($-\muαι$), Armenian ($-m$), Albanian ($-m$)
- formation of *s*-aorists in $*-ah_2s-$ from denominative verbs in $*-ah_2-i\check{e}/o-$: Greek, Armenian and Albanian (see Søbørg 2020: 78–80, 103, elaborating on Klingenschmitt and Matzinger); this connection presupposes that Armenian aorist marker $-c-$ derives from the *s*-aorist
- aorist $*e-k^wle-to$ ‘became’: Greek, Armenian, Albanian (Gr. *ἔπλετο*, Arm. *elew*, OAlb. *cleh*, see LIV² 386–7)
- negation $*(ne) h_2o\check{i}u k^wid$: Gr. *οὐκί*, Arm. *oč’* and Alb. *as* but cf. also, as demonstrated by Fellner (2022), the closely related emphatic negation Toch. A *mā ok*, B *māwk/mā_hk*
- $*ai\check{g}-$ ‘goat’: Greek, Armenian and Albanian
- $*d^heh_1s-$ ‘god’: Gr. *θεός* ‘god’ ($< *d^h_1s-o-$), Arm. *di-k’* ‘(heathen) god’, Phryg. *δεως*
- additional $-a\check{i}(k)-$ in the inflection of the word for ‘woman’: Gr. *γυναῖκ-*, Phryg. acc. *κναικαν*, Alb. *grā* (Matzinger 2000); synchronically, Arm. *kanayk’* is simply the nom.pl. of a stem *kanay-*, but it cannot be excluded that the ending $-k’$ is due to a reinterpretation of a suffixal $-k-$
- $*g^{wh}ermo-$ ‘warm’: a full-grade *mo*-adjective common to Gr. *θερμός*, Arm. *ǰerm* and Alb. *zjarm*

A discussion of the relationship between the Balkan group and Indo-Iranian, including such features as the augment, which may theoretically represent an archaism, is beyond the scope of this chapter.

References

- Adjarian, Hratchia. 1909. *Classification des dialectes arméniens*. Paris: Champion.
- Austin, William M. 1942. Is Armenian an Anatolian language? *Language* 18. 22–5.
- Beekes, Robert S. P. 1988. Laryngeal developments: A survey. In Alfred Bammesberger (ed.), *Die Laryngalthorie und die Rekonstruktion des indogermanischen Laut- und Formensystems*, 59–105. Heidelberg: Winter.

- Beekes, Robert S. P. 2003. Historical phonology of Classical Armenian. In Frederik Kortlandt, *Armeniaca: Comparative notes*, 113–211. Ann Arbor, MI: Caravan Books.
- Bonfante, Giuliano. 1937. Les isoglosses gréco-arméniennes. In Louis Hjelmslev et al. (eds.), *Mélanges linguistiques offerts à Holger Pedersen*, 15–33. Århus: Universitetsforlaget.
- Chantraine, Pierre. 1999. *Dictionnaire étymologique de la langue grec: Histoire des mots*. 2nd ed. Paris: Klincksieck.
- Clackson, James. 1994. *The linguistic relationship between Armenian and Greek*. Oxford: Blackwell.
- Clackson, James. 2005. Review of Kortlandt, *Armeniaca: Comparative notes*, Ann Arbor, MI: Caravan Books, 2003. *Annual of Armenian Linguistics* 24/25. 153–8.
- Clackson, James. 2017. The lexicon of Armenian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1115–32. Berlin: De Gruyter Mouton.
- Demiraj, Bardhyl. 1997. *Albanische Etymologien*. Amsterdam: Rodopi.
- Eichner, Heiner. 1978. Die urindogermanische Wurzel **H₂reu* ‘hell machen’. *Sprache* 24. 144–62.
- Fellner, Hannes A. 2022. Polar life in the Tarim Basin. In Hannes A. Fellner, Melanie Malzahn & Michaël Peyrot (eds.), *Ilyuke wmer ra: Indo-European studies in honor of Georges-Jean Pinault*, 125–37. Ann Arbor, MI: Beech Stave.
- Francis, E. D. 1970. *Greek disyllabic roots: The aorist formations*. PhD thesis, Yale University.
- Gamkrelidze, T. V. 2003. Indo-European and the Glottalic Theory: In defense of ejectives for Proto-Indo-European. In Mary R. Wise et al. (eds.), *Language and life: Essays in memory of Kenneth L. Pike*, 513–31. Dallas, TX: SIL International Publications in Linguistics.
- GEW = Hjalmar Frisk. 1960–72. *Griechisches etymologisches Wörterbuch*. 3 vols. Heidelberg: Winter.
- Godel, Robert. 1975. *An introduction to the study of Classical Armenian*. Wiesbaden: Reichert.
- Greppin, John A. C. 1973. Armenian *h-*, Hittite *h-*, and the Indo-European laryngeal. *Handēs Amsōreay* 87. 61–80.
- Hamp, Eric P. 1976. **g^heiH_o-* ‘live’. In Anna Morpurgo Davies & Wolfgang Meid (eds.), *Studies in Greek, Italic, and Indo-European linguistics offered to Leonard R. Palmer*, 87–91. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- HAB = Hrač’ya Ačařyan. 1971–9. *Hayerēn armatakan bařaran [Armenian root dictionary]*. 4 vols. 2nd ed. (1st ed. 1926–35.). Erevan: Erevani Hamalsaran.
- Holst, Jan Henrik. 2009. *Armenische Studien*. Wiesbaden: Harrassowitz.
- Hyllested, Adam. 2004. Greek *λωτός* and the Indo-European words for ‘blue’. In James Clackson & Birgit Anette Olsen (eds.), *Indo-European word formation*, 59–64. Copenhagen: Museum Tusculanum.
- Ĵahowkryan, Gevorg B. 1987. *Hayoc’ lezvi patmowt’yown: Naxagrayin žamanakašrjan [History of the Armenian language: The pre-literary period]*. Erevan: Haykakan SSH GA Hratarakč’ut’yun.
- Jasanoff, Jay H. 1979. Notes on the Armenian personal endings. *Zeitschrift für vergleichende Sprachforschung* 93. 133–49.

- Job, Michael. 1995. Zum Lautwandel im Armenischen: Probleme einer relativen Chronologie. In Wojciech Smoczyński (ed.), *Kuryłowicz memorial volume*. Vol. 1, 291–311. Cracow: Universitas.
- Kim, Ronald I. 2018. Greco-Armenian. The persistence of a myth. *Indogermanische Forschungen* 123. 247–71.
- Klein, Jared S. 2007. Classical Armenian morphology. In Alan S. Kaye (ed.), *Morphologies of Asia and Africa*. Vol. 2, 1051–86. Winona Lake, IN: Eisenbauns.
- Klingenschmitt, Gert. 1970. Griechisch *ἰλάσκεισθαι*. *Münchener Studien zur Sprachwissenschaft* 28. 75–88.
- Klingenschmitt, Gert. 1982. *Das altarmenische Verbum*. Wiesbaden: Reichert.
- Klingenschmitt, Gert. 1994. Die Verwandtschaftsverhältnisse der indogermanischen Sprachen. In Jens Elmegård Rasmussen (ed.), *In honorem Holger Pedersen*, 235–51. Wiesbaden: Reichert.
- Kocharov, Petr. 2019. *Old Armenian nasal verbs: Archaisms and innovations*. PhD thesis, Leiden University.
- Kortlandt, Frederik. 1975. A note on the Armenian palatalization. *Zeitschrift für vergleichende Sprachforschung* 89. 43–5.
- Kortlandt, Frederik. 1980a. On the relative chronology of Armenian sound changes. In John A. C. Greppin (ed.), *Proceedings of the First International Conference on Armenian Linguistics*, 97–106. Delmar, NY: Caravan Books.
- Kortlandt, Frederik. 1980b. *H₂O* and *oH₂*. *Lingua Posnaniensis* 23. 127–8.
- Kortlandt, Frederik. 1984. Proto-Armenian case endings. In Gevorg B. Ĵahowkjan & A. Pisowicz (eds.), *International Symposium on Armenian Linguistics, Erevan, September 21–25, 1982: Reports*, 97–106. Yerevan: Armenian Academy of Sciences.
- Kortlandt, Frederik. 1986. Armenian and Albanian. In Maurice Léroy & Francine Mawet (eds.), *La place de l'arménien dans les langues indo-européennes*, 38–47. Leuven: Peeters.
- Kortlandt, Frederik. 1987. Notes on Armenian historical phonology V. *Studia Caucasica* 7. 61–5.
- Kristoffersen, Tore Røvs. 2019. *Monophthongization of u-diphthong before labial consonant in Greek*. Master's thesis, University of Copenhagen.
- Kroonen, Guus. 2012. Non-Indo-European root nouns in Germanic: Evidence in support of the Agricultural Substrate Hypothesis. In Riho Grünthal & Petri Kallio (eds.), *A linguistic map of prehistoric northern Europe*, 239–60. Helsinki: Suomalais-Ugrilainen Seura.
- Kümmel, Martin. 2012. Lautgeschichte und Etymologie: Jungavestisch *mas-* 'groß'. Presentation at the conference *Iranian and Indo-European Onomastics and Linguistics: Symposium in memory of Manfred Mayrhofer (1926–2011)*, Vienna, 10 May 2012.
- Kümmel, Martin. 2017. Even more traces of the accent in Armenian? The development of tenues after sonorants. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Usque ad radices: Indo-European studies in honour of Birgit Anette Olsen*, 439–52. Copenhagen: Museum Tusculanum.
- Lamberterie, Charles de. 1982. Review of Schmitt 1981. *Bulletin de la Société de Linguistique de Paris* 77. 80–6.
- Lamberterie, Charles de. 1983. Une bagatelle étymologique (*artewanunk* 'δρεπάνη). *Revue des études arméniennes* n.s. 17. 21–2.

- Lamberterie, Charles de. 1989. Introduction à l'arménien classique. *LALIES* 10. 233–89.
- Lamberterie, Charles de. 1990. *Les adjectifs grecs en -vç*. 2 vols. Leuven: Peeters.
- Lamberterie, Charles de. 2013. Grec, phrygien, arménien: des anciens aux modernes. *Journal des savants* 2013(1). 3–69.
- Lindeman, Fredrik Otto. 1987. *Introduction to the “Laryngeal Theory”*. Oslo: Norwegian University Press.
- LIV² = Helmut Rix et al. (eds.). 2001. *Lexikon der indogermanischen Verben*. 2nd ed. Wiesbaden: Reichert.
- Lubotsky, Alexander. 2002. The Indo-Iranian word for ‘shank, shin’. *Journal of the American Oriental Society* 122. 318–24.
- Macak, Martin. 2017. The phonology of Classical Armenian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1037–79. Berlin: De Gruyter Mouton.
- Martirosyan, Hrach K. 2010. *Etymological dictionary of the Armenian inherited lexicon*. Leiden: Brill.
- Martirosyan, Hrach K. 2013. The place of Armenian in the Indo-European language family: The relationship with Greek and Indo-Iranian. *Journal of Language Relationship* 10. 85–137.
- Martirosyan, Hrach K. 2018. The Armenian dialects. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia: An areal perspective*, 46–105. Berlin: De Gruyter Mouton.
- Martzloff, Vincent. 2016. Le verbe arménien *gerem* ‘emmener en captivité’ et le traitement phonétique de **d^h* intervocalique en arménien. *Wék^{vos}* 2. 109–77.
- Matzinger, Joachim. 1992. *Der armenische Reflex von idg. **z_u**. Vienna.
- Matzinger, Joachim. 2000. Albanisch *grā* ‘Frauen’. *Münchener Studien zur Sprachwissenschaft* 60. 75–87.
- Matzinger, Joachim. 2005a. *Untersuchungen zum altarmenischen Nomen: Die Flexion des Substantivs*. Dettelbach: Röhl.
- Matzinger, Joachim. 2005b. Phrygisch und Armenisch. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 375–94. Wiesbaden: Reichert.
- Matzinger, Joachim. 2012. “Zwischensprachen” – Areallinguistische Bemerkungen aus dem Bereich des Balkanindogermanischen. In Velizar Sadovski & David Stifter (eds.), *Iranistische und indogermanischen Beiträge in memoriam Jochem Schindler (1944–1994)*, 137–59. Vienna: Österreichische Akademie der Wissenschaften.
- Meillet, Antoine. 1904. Remarques sur la grammaire historique de l'arménien de Cilicie, de M. J. Karst. *Zeitschrift für armenische Philologie* 2. 18–28.
- Meillet, Antoine. 1936. *Esquisse d'une grammaire compare de l'arménien classique*. Vienna: Mekhitharistes.
- Meissner, Torsten. 2006. *S-stem nouns and adjectives in Greek and Proto-Indo-European: A diachronic study in word formation*. New York: Oxford University Press.
- Meyer, Robin. 2017. *Iranian-Armenian language contact in and before the 5th century CE: An investigation into pattern replication and societal multilingualism*. PhD thesis, University of Oxford.
- Normier, Rudolf. 1977. Idg. Konsonantismus, germ. “Lautverschiebung” und Vernersches Gesetz. *Zeitschrift für vergleichende Sprachforschung* 91. 171–218.

- Olsen, Birgit A. 1985. On the development of Indo-European prothetic vowels in Armenian. *Revue des études arméniennes* 19. 5–17.
- Olsen, Birgit A. 1986. The Armenian continuations of Indo-European intervocalic *w. *Annual of Armenian Linguistics* 7. 51–6.
- Olsen, Birgit A. 1988. On the formation of the subjunctive paradigm in Classical Armenian. *Annual of Armenian Linguistics* 10. 5–25.
- Olsen, Birgit A. 1989. A trace of Indo-European accent in Armenian. *Historische Sprachforschung* 102. 220–40.
- Olsen, Birgit A. 1992. The development of high vowel plus laryngeal in Armenian. In John A. C. Greppin (ed.), *Proceedings of the First International Conference on Armenian Linguistics*, 129–46. Delmar, NY: Caravan Books.
- Olsen, Birgit A. 1993. On the origins of Armenian c'. In L. Hovsepian, N. Parnassian & S. Simonian (eds.), *Proceedings of the Second International Symposium on Armenian Linguistics*, 25–48. Yerevan: Armenian Academy Press.
- Olsen, Birgit A. 1999. *The noun in biblical Armenian*. Berlin: Mouton de Gruyter.
- Olsen, Birgit A. 2009. The conditioning of laryngeal breaking in Greek. In Rosemarie Lühr and Sabine Ziegler (eds.), *Protolanguage and prehistory*, 348–65. Wiesbaden: Reichert.
- Olsen, Birgit A. 2017a. The morphology of Armenian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1080–97. Berlin: De Gruyter Mouton.
- Olsen, Birgit A. 2017b. Armenian. In Mate Kapović (ed.), *The Indo-European languages*. 2nd ed., 421–51. London: Routledge.
- Olsen, Birgit A. 2019. Aspects of family structure among the Indo-Europeans. In Birgit A. Olsen, Thomas Olander & Kristian Kristiansen (eds.), *Tracing the Indo-Europeans: New evidence from archaeology and historical linguistics*, 145–63. Oxford: Oxbow.
- Olsen, Birgit A. 2020. An Armenian sound law revisited: *-eu- > -iw-? In Claire Le Feuvre & Daniel Petit (eds.), *Ονομάτων Ἱστορία: Mélanges offerts à Charles de Lamberterie*, 115–24. Leuven: Peeters.
- Orongo, Alessandro. 2017. The documentation of Armenian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1028–37. Berlin: De Gruyter Mouton.
- Pedersen, Holger. 1905. Zur armenischen sprachgeschichte. *Zeitschrift für vergleichende Sprachforschung* 38. 194–240.
- Pedersen, Holger. 1924. Armenier. In M. Ebert (ed.), *Reallexikon der Vorgeschichte*, 219–26. Berlin: de Gruyter.
- Pisani, Vittore. 1978. Studi sulla fonetica dell'armeno. In Vittore Pisani (ed.), *Mantissa*, 255–317. Brescia: Paideia.
- Rasmussen, Jens Elmegård. 1991. *ih, uh and Rh* in Indo-European: A phonetic interpretation. *Copenhagen Working Papers in Linguistics* 1. 127–39.
- Ravnæs, Erling. 1991. *The chronology of the sound changes from Proto-Indo-European to Classical Armenian*. PhD Thesis, University of Oslo.
- Schmitt, Rüdiger. 1981. *Grammatik des Klassisch-Armenischen mit sprachvergleichenden Erläuterungen*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Solta, Georg Renatus. 1960. *Die Stellung des Armenischen im Kreise der indogermanischen Sprachen*. Vienna: Mechitaristen-Buchdruckerei.

- Solta, Georg Renatus. 1963. Die armenische Sprache. In Bertold Spuler (ed.), *Handbuch der Orientalistik*. Vol. 1. Part 7. *Armenisch und kaukasische Sprachen*, 80–128. Leiden: Brill.
- Stempel, Reinhard. 1983. *Die infiniten Verbalformen des Armenischen*. Frankfurt am Main: Lang.
- Søborg, Tobias Mosbæk. 2020. *Sigmatic verbal formations in Anatolian and Indo-European: A cladistic study*. PhD thesis, University of Copenhagen.
- Taillardat, J. 1977. Images et matrices métaphoriques. *Bulletin de l'Association Guillaume Budé* 36. 334–54.
- Weiss, Michael. 2020. *Outline of the historical and comparative grammar of Latin*. 2nd ed. Ann Arbor, MI: Beech Stave.
- Weitenberg, Jos. J. S. 2017. The dialectology of Armenian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 2, 1132–46. Berlin: De Gruyter Mouton.
- Winter, Werner. 1965. Armenian evidence. In Werner Winter (ed.), *Evidence for laryngeals*, 100–15. London: Mouton.
- Witczak, Krzysztof T. 1991. Armenian *op'i* 'white poplar, *Populus alba* L.' and the development of **ps* in Armenian. *Annual of Armenian Linguistics* 12. 65–75.
- Woodhouse, Robert. 2015. Some observations on the putative dual reflexes of PIE **CRHC* in Greek and Armenian, Francis' law and Greek *αἰχμήν* 'neck, etc.' *Studia Etymologica Cracoviensia* 4. 57–272.

13 Albanian

Adam Hyllested & Brian D. Joseph

13.1 Introduction

Albanian is sometimes considered the stepchild of Indo-European linguistics, for various reasons. For one, it is the latest attested IE branch; its first documentation is a 1462 one-line baptismal formula, and the first substantial text the 1555 *Missal* of Gjon Buzuku. Due to this late attestation, many details of its historical development are shrouded in mystery, and its present form does not always appear obviously Indo-European. Consider, for example, the numerals *gjashtë* ‘6’ and *tetë* ‘8’, which despite looking strikingly different from, say, Latin *sex* and *octō*, in fact reflect the expected outcomes of PIE **séks-tV-* and **óktō-tV-*.

Moreover, the complicating factor of heavy external influence can make it difficult to determine what is inherited from PIE. Not only are there Albanian borrowings from Ancient Greek, Latin (*sensu lato*), Slavic, Turkish, and Italian, as well as from neighbouring Balkan languages, but there is also structural convergence with other Balkan languages, especially Modern Greek, Macedonian, Aromanian, and Romani, but also Turkish, and, by extension, Bulgarian, Meglenoromanian, and Romanian. This convergence covers phonology, e.g. voicing of nasal + stop clusters, as in *këndoj* ‘sing’ (borrowed from Latin *cantō*), matching a development in Greek and Aromanian; morphology, e.g. the merger of genitive and dative cases, matching a development in Greek, Aromanian, Romanian, Macedonian, and Bulgarian; syntax, e.g. doubling of direct or indirect objects by weak pronouns, matching a development in Greek, Aromanian, Romanian, Macedonian, Bulgarian, and to some extent, Romani; and semantics, e.g. creation of admirative mood forms to mark non-confirmativity, matching a development in Macedonian, Bulgarian, and Turkish.

13.2 Evidence for the Albanian Branch

These difficulties notwithstanding, several innovations define Albanian and set it apart from all other branches of IE, including

- **s > [j]* (in IPA, spelled ⟨gj⟩ in standard Albanian orthography) in initial position before a stressed vowel, cf. *gjashtë* ‘6’ < **séks-tV-* vs. *shtatë* ‘7’ < **septm̥-tV-*. ⟨gj⟩

represents a voiced dorsopalatal stop, though with varied secondary outcomes dialectally. This change is unparalleled within IE.

- $*k > [\theta]$ (spelled ⟨th⟩), a change found only also in Old Persian among other IE branches; e.g. *athēt* ‘harsh, sour’ < $*ak̑$ ‘sharp’ (cf. Ved. *ás-man-* ‘stone’)
- $*g^{(h)} > [\delta]$ (spelled ⟨dh⟩), also unparalleled within IE,¹ e.g. *udhë* ‘way’ < $*uḡ^{h-o-}$ (the root of Lat. *veh-ō* ‘convey’)
- loss of word-internal voiced stops under certain conditions, e.g. *ujë* ‘water’ < PALb. $*ud-r-jā$
- $*ō > e$, as in *tetë* ‘8’ < $*ōktō-tV-$
- $*ē > o$, as in *mos* ‘not; don’t!; lest’ < $*meh_1-k^{w}id$ (cf. Gr. *μή*)
- *-ni* as 2pl. non-past verbal ending, e.g. present indicative *ke-ni* ‘you all have’, imperative *ki-ni* ‘you all have!’, from a reanalysed and repurposed adverbial $*nū$ ‘now’ (Rasmussen 1985)
- a postposed definite article, as in *det-i* ‘the sea’ (literally ‘sea-the’).²

These characteristics give ample cause for treating Albanian as a separate branch within IE, even with various complications in analysing forms.

13.3 The Internal Structure of Albanian

Despite constituting its own branch within IE, Albanian is hardly a linguistic monolith. In fact, there are major dialect divisions within the branch, the oldest and most important being a north–south one: the Geg dialect group occurs north of the Shkumbin river (roughly in the middle of present-day Albania), thus covering northern Albanian and the Albanian of the nation-states of North Macedonia, Kosova, and Montenegro, while the Tosk group occurs south of the river, and includes the Arbëresh diaspora communities of southern Italy and the Arvanitika diaspora communities scattered around Greece.

Dialect differences separating Geg and Tosk involve all levels of linguistic structure. In phonology, Geg has nasalized vowels whereas Tosk has lost nasalization (e.g. *âsht* ‘is’ vs. Tosk *është* < $*ensti$ < PIE $*h_1en-h_1esti$), maintains intervocalic *-n-* whereas Tosk denasalizes it to *-r-* (e.g. *venë* ‘wine’ vs. Tosk *verë*) and has reduced nasal-plus-stop clusters to nasals whereas Tosk maintains the clusters (e.g. *nimoj* ‘I-help’ vs. Tosk *ndihmoj*). In morphology, Geg has participials in *-m-* (among other endings) whereas Tosk mostly uses *-uar* (e.g. *harrum* ‘forgotten’ vs. Tosk *harruar*), and Geg forms its future tense with an

¹ The notation $\hat{g}^{(h)}$ indicates that the PIE voiced aspirated and voiced plain stops generally merged in Albanian; while this development is characteristic of Albanian, it is not particularly striking within IE, occurring, presumably independently, in Anatolian, Balto-Slavic, Celtic, Iranian, and Tocharian.

² This feature is found also in neighbouring languages, especially Aromanian, Macedonian, and Romanian, suggesting causality through contact rather than internal innovation within Albanian. However, Hamp 1982 argues that the ancient toponym *Drobeta* (in present-day Romania) reflects a Roman misinterpretation of $*druwā-tā$ ‘the wooded (place)’, with a postposed definite article, suggesting it reflects an old Albanian syntagm.

inflected form of ‘have’ plus an infinitive (consisting of *me* with a participial) whereas Tosk uses an invariant (3sg.) form of ‘want’ with an inflected subjunctive with the modal marker *të* (e.g. *ke me shkue* ‘you will go’ (literally “you-have to gone”) vs. Tosk *do të shkosh* (“it-wants that you-go”)). In syntax, Geg uses its (uninflected) infinitive with *me* in complement structures where Tosk uses the (inflected) subjunctive with *të*, e.g. *filloj me shkue* ‘I begin to go’ (literally “I-begin to gone”) vs. Tosk *filloj të shkoj* (literally “I-begin that I-go”). Finally, there are lexical differences, e.g. Geg *tamël* ‘milk’ vs. Tosk *qumësht*.

Within the Geg and the Tosk dialect complexes, there is much regional variation, the details of which are beyond the scope of this chapter. It can be noted, though, that diaspora varieties of Tosk show the effects of differential contact situations: Arbëresh in Italy not only has many Italian loans not found in Balkan Tosk, e.g. *kamineta* ‘chimney’ (cf. Italian *camineta* ‘fireplace’) but also lacks Turkish loanwords (cf. Balkan Tosk *oxhak* ‘chimney, fireplace’, from Turkish *ocak*), reflecting its absence from the Balkans after approximately the fifteenth century. Similarly, Arvanitika in Greece shows various Greek features not generally found in Tosk; for instance, according to Sandfeld (1930: 104), in Arvanitika, *mnj* (Sandfeld’s notation) occurs for *mj* elsewhere in Balkan Tosk, e.g. *mnjekrë* ‘chin; beard’ (vs. general Tosk *mjekër*), a shift he states is “comme en grec” (cf. Thumb 1912: §30, who reports colloquial Greek *μνιά* ‘one.FEM’ (presumably [mɲja] or [mɲa]) versus earlier, and still occurring, *μιά* ([mjá])).

13.4 The Relationship of Albanian to the Other Branches

Albanian shows mixed dialectal affinities, sharing key features with different sets of languages within IE. This situation makes for a complicated determination of how to subgroup Albanian with other branches. Ultimately, although no consensus prevails as to the exact classification of Albanian, we argue here that lexical and morphological isoglosses point to a Greek-Albanian subgroup, a grouping suggested by computational phylogenetic methodology in Chang et al. 2015 (see Section 13.5.2; note also Holm 2011).

We base our discussion largely on significant, non-trivial innovations Albanian shares with other branches. However, what counts as a shared innovation as opposed to a shared retention of course depends on decisions made about the nature of the proto-language in question. Thus, assessments about subgrouping can become complicated and involved.

For instance,³ Cowgill (1960) proposed that Greek *οὐ(κί)* ‘not’ could be connected with Armenian *ոչ* ‘not’, with both deriving from a phrase **ne ...*

³ Other cases like this of what we consider retentions, but which some scholars might see as innovations, are the use in prohibitions of **meh* (Alb. *mos*, Gr. *μὴ*; see also Section 13.4.7 Inflection and Morphosyntax) and the use of the augment in marking past tense forms. Space limitations preclude discussion here; see Joseph 2013.

h₂óju kʷid, composed of the negative marker **ne*, the noun **h₂óju* ‘life-force’, and the indefinite pronoun **kʷid*, thus originally “not on (your) life; not at all”, as an emphatic negator. He conjectured, following Pedersen 1900, that the Albanian negative *as* ‘nor, and not’ might belong here too but was reluctant to pursue the connection. Joseph (2005; 2022) has followed up on the Albanian angle, arguing that the negative prefix *as-* ‘not’, as in *as-gjë* ‘nothing’ (cf. *gjë* ‘thing’), is what matches *ov(ki)* and *oč*.⁴ Ostensibly, this **ne* ... *h₂óju kʷid* phrasal negation could be a shared innovation linking Albanian, Armenian, and Greek (Section 13.4.8), if restricted to those branches. However, Garnier 2014 and Fellner 2022 have argued that Latin *haud* ‘not’ and Toch.A *mā ok*, Toch.B *mawk*, *maṇk*, respectively, also reflect *(*ne*) ... *h₂óju kʷid*, so this negator is shared by languages that do not otherwise show evidence for being subgrouped together. Thus **ne* ... *h₂óju kʷid* must be of PIE age, so its occurrence in these languages is a shared retention inherited in each and therefore irrelevant to subgrouping. Any potential shared innovation in principle must be examined carefully to determine its status vis-à-vis innovation versus retention.

As noted above, there are numerous, often contradictory, indications of close connections between Albanian and other branches of IE, and though we ultimately favour the connection with Greek, we review here the evidence that aligns Albanian with one or another branch of IE.

13.4.1 Albanian and Balto-Slavic

Various features connect Albanian with Balto-Slavic. We mention a few here, and point interested readers to Porzig 1954: 174–7, Jokl 1963, Čabej 1975, Huld 1984: 166, Orel 1994; 2000: 254–6 for further details and assessment.

13.4.1.1 -teen Numerals Albanian forms the teen numerals eleven to nineteen using a pattern of DIGIT-on-TEN, e.g. *njëmbëdhjetë* ‘eleven’ (cf. *një* ‘one’, *mbi* ‘on’, *dhjetë* ‘ten’), that seems to parallel Slavic (e.g. Ru. *odinnadcat* ‘eleven’ (cf. *odin* ‘one’, *na* ‘on’, *désjat* ‘ten’)) and part of Baltic, specifically Latvian (e.g. *vienpadsmit* ‘eleven’; Lithuanian aligns with Germanic here, using a formative based on **leikʷ-* ‘leave’, not a form of ‘ten’). However, there is one key difference between the Albanian and the Slavic/Latvian patterns. Albanian, along with Romanian, has a feminine form of ‘ten’, shown by the use of the feminine *tri* ‘three’ with *dhjetë* ‘ten’ in the formation of ‘thirty’, *tridhjetë*, whereas Slavic has a masculine form, as in the Russian use

⁴ The relationship between the free word *as* and the prefix *as-* is disputed; Joseph sees them as having different origins, while others connect them. That issue is irrelevant here, as the fact of there being some Albanian cognate to the Greek and Armenian forms is all that matters in this case. See also Hackstein 2020 on sources of negation markers in Albanian, including **ne* ... *h₂óju kʷid*.

of masculine *dva* ‘two’ in the formation of ‘twenty’, *dvádecát* (literally “two tens”); Romanian for ‘twenty’ is *douăzeci* ‘twenty’ (literally “two tens”), with feminine *două*, thus with feminine ‘ten’.

Following [Hamp \(1992\)](#), these facts can be interpreted for the Balkans as follows. The variety of IE destined to become Albanian (Hamp’s “Albanoid”) was a Northern IE language, grouped with or in contact with Germanic and Balto-Slavic. Within Baltic, Lithuanian absorbed the teen-numeral pattern of Germanic, whereas Latvian interacted with Slavic and Albanoid, an inner-Baltic difference that makes sense geographically. Albanoid, along with Latvian and Proto-Slavic, developed the DIGIT-on-TEN pattern, presumably an innovation in one language that spread by contact into the others, but its speakers changed this pattern as they moved south into the Balkans and came into contact with the variety of Latin that some of its speakers shifted to, yielding Romanian. This scenario accounts for both the similarities between Albanian and Slavic (and Latvian) and the differences within Baltic, while still allowing for the specific Albanian–Romanian parallel to emerge.

13.4.1.2 Winter’s Law [Winter \(1978\)](#) posited for Baltic and Slavic the lengthening of vowels before PIE voiced plain stops (mediae, e.g. **d*), a prime example being Balto-Slavic **sēd-* ‘sit’ (cf. infinitives Lith. *sėsti* and OCS *sěsti*), from PIE **sed-*. Albanian seems to similarly show this development, in forms such as *rronj* ‘endure’ < **rēg-n-* (with *o* regularly from earlier **ē*; for the root, cf. Gr. *ὀρέγω* ‘extend’) or *erë* ‘smell’ < **ōd-r-* (PIE **h₃ed-*, cf. Lat. *odor*), although this may alternatively reflect compensatory lengthening with the loss of the stop ([Hyllested 2013](#)).

13.4.1.3 Lexical Isoglosses Several scholars have noted sizeable lexical overlap between Balto-Slavic and Albanian. [Orel \(1998: 250–6\)](#) counts twenty-four shared items, deeming this group of isoglosses the “most important and significant” one. As many as forty-eight words are allegedly shared between Albanian and Baltic only, leading Orel to call this connection “particularly close”, while he further lists twenty-two terms shared just by Albanian and Slavic (“not as frequent as Baltic ones”).

However, not all of these etymologies appear equally convincing. For example, Alb. *bac* ‘elder brother; uncle’ must be borrowed from Slav. **bat’a* ‘elder brother; father’, not cognate with it ([Hyllested 2020: 402](#)); Alb. *shtrep*, *shtrebë* ‘cheese-fly larva’, rather than being related to Slav. **strupъ* ‘scab’, belongs with Gr. *σπρέφω* ‘twist’, as is not least apparent from its inner-Albanian cognate *shtrembet* ‘be crooked’ ([Hyllested 2016: 75](#)); and Alb. *murg* ‘dark, grey’ ~ Lith. *márgas* ‘colorful’ do not constitute an isogloss but are clearly related to both PGmc. **murkaz* ‘dark’, Gr. *ἀμωρός* ‘dark’ and Slav. **mergъ* ‘brown’.

Crucially, the more promising of these *comparanda* are, in most cases, morphologically and/or semantically more distant from each other than the proposed Helleno-Albanian isoglosses. Alb. *brez* ‘belt’ vs. Lith. *briaunà* ‘edge’ is a typical example: these two words undoubtedly contain the same IE root but with markedly different word-formation and meanings that differ significantly. Thus, while the item is useful in a general comparative analysis, it is less so as evidence for subgrouping. A systematic analysis of all relevant forms goes beyond our scope, but one can fairly say that the number of closely knit lexemes with strong etymologies is in fact not significantly higher between Albanian and Balto-Slavic than one would expect between any two IE branches.

13.4.2 Albanian and Armenian

Considering the large number of shared innovations between Albanian and Greek on the one hand (Section 13.4.7) and between Greek and Armenian on the other (Section 12.4.1), it is perhaps surprising how few can be found between Albanian and Armenian only. This does not speak against a Palaeo-Balkan subgroup encompassing all three since it may simply reflect the fact that Greek preserves so much more IE lexical material, including Balkanic innovations, than the other two.⁵ Most famous among the relevant isoglosses is Alb. *zog* ‘bird; nestling; (dial.) animal young’ ~ Arm. *jag* ‘little bird, sparrow; nestling’, as if from a protoform **ǵʰuāǵʰu-* (Jokl 1963: 152; Olsen 1999: 110–11); however, it may constitute a shared retention since its root etymology is unknown.

A shared inflectional feature is the new masculine **smi-ī-* for the numeral ‘one’, Alb. *një* and Arm. *mi*, based on the Balkanic feminine **smi-ī-a* with breaking from PIE **sm-ih₂* as in Gr. *μία* (Klingenschmitt n.d.: 22).

In derivational morphology, Armenian and Albanian share a productive agent-noun suffix **-ikʷiō-* > Arm. *-ičʻ*, Alb. *-ës* (Matzinger 2016: 167; Thorsø 2019: 252), which we see as derived from PIE **kʷei-* ‘gather’ (cf. Gr. *ποιέω* ‘make’).

One phonological development shared by Albanian and Armenian is loss of **m* in the cluster **-ms-*, cf. Alb. *mish* ‘meat’ ~ Arm. *mis* ‘id.’ < PIE **mems-o-*; Arm. *ows* ‘shoulder’ vs. Gr. *ὤμος* ‘shoulder’ < PIE **h₁ómsos*. This must however reflect two parallel developments if, as we argue, Albanian and Greek (or, for that matter, Armenian and Greek) form a subgroup within Balkanic, since Greek preserves the **-m-*.

Other joint phonological features relate to *centum*–*satem* behavior and are mostly systematically parallel, not necessarily substantially identical. First and

⁵ See Section 13.4.8 on innovations shared by the entire proposed Balkan group.

foremost, like Albanian, Armenian keeps a three-way distinction of PIE dorsals (see Section 13.5.1). But both languages also have a development of PIE $*k_u$ - and $*g^h_u$ -, which, like everywhere in the *satem* area proper, is different from both that of the palatals and that of labiovelars but at the same time, unlike Indo-Iranian and Balto-Slavic, shows no direct trace of the semivowel; e.g. Alb. *zë*, def. *zëri* (Geg *zâ*, *zâni*) ‘voice’, Arm. *jayn* ‘voice, sound’ ~ OCS *zvонѣ* ‘noise’ < PIE $*g^h_uónos$.

13.4.3 Albanian and Celtic

Few traits, almost exclusively lexical in nature, link Albanian specifically with Celtic. A quite optimistic pioneering collection of isoglosses by Jokl 1927 was subjected to critical scrutiny by Çabej 1969, who effectively disqualified much of the evidence. Most famous is the similarity between Alb. *gju* ‘knee’, S Tosk *glu*, Geg *gju*, def. *gjuni*, ~ PCelt. $*glunos$ ‘knee’ (OIr. *glún*, Welsh *glin*), apparently involving a new stem-form $*gnu-n-$ from PIE $*g^henu$ with subsequent dissimilation to $*glu-n-$.

The remaining evidence amounts to nothing more than what would be expected statistically; Orel (2000) mentions only six items. Moreover, the picture is somewhat blurred by the fact that many apparent shared lexemes are likely early Celtic borrowings into Proto-Albanian from when Celtic tribes such as the Serdi and the Scordisci settled in the Balkans in the third century BCE. This may, e.g., be the case with Alb. *shqipe* ‘eagle’, which, like Welsh *ysglyn* ‘eagle’, is derivable from a proto-form $*sklubo-$, metathesized from earlier $*skublo-$ from which the other attested Celtic forms developed (Hyllested 2016: 76–7).

13.4.4 Albanian and Germanic

Ringe, Warnow, and Taylor (2002), in a statistical-quantitative analysis of the IE lexicon, reached the apparent result of an Albanian subgroup with Germanic, the significance of which the authors themselves downplayed, and with good reason: the absolute number of lexical cognates shared by these branches only is relatively moderate. Orel 1998: 253–4 lists just thirteen, not all with equally valid etymologies; for example, *tym* ‘smoke’ must be borrowed from Gr. *θυμός* (with an older meaning than the attested ‘anger’), rather than related to PGmc. $*ēðumaz$ ‘breath’.⁶ Moreover, the lexical isoglosses are not corroborated by many shared grammatical elements or features.

⁶ One oft-mentioned item is Alb. *det* ‘sea’, Arbëresh *dej(ë)t*, usually etymologized as PAIb. $*deubeta$, corresponding to PGmc. $*deupīþō-$ ‘depth’. Hyllested (2016: 71 n. 12) instead suggests it could be a borrowing from Gr. *δέλτα* ‘river delta’. At least two other Albanian

There are nonetheless some remarkable cases of shared word-formation. One recently published etymology is *hundë* ‘nose’ < PALb. **skuntā* ~ Far., SWNw. *skon* ‘snout’ < PGmc. **skuna-* (Hyllested 2012). Alb. *delme* ‘sheep’ is only a metathesis away from corresponding regularly to Dalecarlian *tembel* ‘sheep’ < PGmc. **tamila-*, a derivative of PGmc. **tamjan* ‘to tame’ < PIE **demH-*; treating the nasal rather than the lateral as original to the Albanian root is supported by the synchronically suppletive plural *dhëndë* < **domH-it-eh₂*, literally ‘the tamed (collective of animals)’.

13.4.5 Albanian and Italic

As stated by Huld (1984: 168): “Relations between Albanian and Italic are largely negligible”. Most prominent among the vanishingly few shared innovations is the lexical pair Alb. *bir* ‘son’, *bijë* ‘daughter’ (as well as Mess. *bilia* ‘daughter’), which is likely identical to Lat. *filius*, *filia*, respectively (Hyllested 2020: 421–2). Albanian *hi*, Geg *hî*, def. *hîni* ‘ashes’ < **sken-is-* seems to agree in ablaut with Lat. *cinis* ‘cold ashes’ < **ken-is-* vs. Gr. *κόινς* ‘dust; ashes’ and Toch.B *kentse* ‘rust’ < **koniso-*, but both forms are probably old in IE, and the equation with Albanian is far from certain anyway (Hyllested 2012: 76 n. 4).

13.4.6 Albanian and Indo-Iranian

Jokl (1963: 152), in his somewhat inconclusive posthumous work, listed eight lexical parallels between Albanian and Indo-Iranian, almost none of which, however, constitute exclusive isoglosses, as Jokl himself acknowledged. Even his flagship first item, Alb. *dhëndër(r)*, Geg *dhândër(r)* ‘son-in-law; bridegroom’, which on the surface looks like the same **-ter* formation from PIE **ĝem(H)-* as Ved. *jāmātar-*, YAv. *zāmātar-* ‘son-in-law’, may simply owe its *-d-* to inner-Albanian epenthesis as in the rhyming word *ëndër(r)* ‘dream’ < PIE **Hon-r-jo-*, while Indo-Iranian **-tar* can be analogical from other kinship terms. In that case, Albanian formally agrees with Lat. *gener* and Gr. *γαμβρός* instead.⁷

Orel’s (2000: 260) more recent list of ten items suffers from the same conspicuous weaknesses; for example, Alb. *thadër* ‘double-sided axe’ does not actually form a unique isogloss with Ved. Br.+ *śāstra-* ‘knife; sword’, since Lat. *castrum* ‘knife’ represents an identical formation < PIE **kas-trom*, lit.

words from the same semantic field are Greek borrowings: *pellg* ‘pond; basin; depth’ ⇐ *πέλαγος* ‘sea’ and *zall* ‘riverbank, river sand’ ⇐ *αἰγιαλός* ‘sea-shore’.

⁷ The irregular and unparallelled plural *dhëndûrë*, North Geg *dhândorrë* is probably due to later conflation with Lat. *genitôres* ‘begetters’ (i.e., of heirs, cf. Eng. *beget an heir*), where the significant position of the plural must be seen in the light of traditional Balkan household structures with several married couples under one roof.

‘cutting-instrument’. A critical assessment of some further oft-mentioned items is provided by Huld (1984: 167).

13.4.7 Albanian and Greek

As noted above, our ultimate assessment treats Albanian and Greek as particularly close relatives within Indo-European. We find the number of innovations shared only by Albanian and Greek to be overwhelming, thus pointing compellingly to a Helleno-Albanian subgroup. In this section, we offer an overview of shared developments, without claiming exhaustiveness. The evidence is mostly morphological and lexical in nature, involving particular lexical items or details of word-formation, but there are also several phonological commonalities.⁸

13.4.7.1 Phonology

1. Initial **j-* has a twofold reflex in both languages: (a) an obstruent **dz-* > Alb. *gj-*, Gr. ζ-, which already appears in Mycenaean, vs. (b) a preserved **j-* > Alb. *j-*, PGr. **j-*, which later yielded *h-* in early Greek, but is still partially retained in Mycenaean. For Greek, the conditioning is famously disputed.⁹ Despite the fact that a similar double reflex between *j-* and *gj-* has long been recognized in Albanian,¹⁰ it has hitherto gone unnoticed that the distribution between individual lexemes is identical in both languages: Alb. *n-gjesh* ‘knead’ (< **iós-(i)je-*) ~ ζέω ‘boil, seethe’ < **jes-* ‘boil; ferment’; Alb. *gjesh* ‘gird’ ~ Gr. ζώννυμι ‘id.’ < PIE **ieh₃s-*; Arbëresh *gjër* ‘soup’, Geg *gjânë* ‘silt, mudbed’ < **iouh₃-(m)n-o-* ~ Gr. ζῆμη ‘sourdough’, ζωμός ‘sauce; broth’ < **ieuh₃-s-* ‘mix sth. moist’; vs. Alb. *ju* ‘you (2pl.)’ ~ Gr. ὅμεις ‘id.’ (although the latter may instead continue PIE acc. **us-mé*); Alb. *a-jo* ‘she’ ~ Gr. rel. pron. f. ἡ < **ieh₂*; and Alb. *josh* ‘fondle, caress’ < **ieud^h-s-* (cf. for

⁸ Space does not allow a word-by-word treatment of purported isoglosses whose validity for various reasons we reject. A few examples may illustrate: Alb. *egër* ‘wild’ must be borrowed from Gr. ἄγριος ‘id.’, not a cognate, since the PIE root has **-g-*, which yields Alb. *dh*. The singularized plural *dhemje* ‘caterpillar; maggot’ is unrelated to Gr. δειλέας ‘leech’; the variant *vemje* shows it is instead a borrowing from the Slavic collective noun **vьrmьje* ‘insects and worms’ with regular development of *v-* > *dh-* / *_VCC* where one consonant is a labial. And while Alb. *derr* ‘pig’ ~ Gr. χοίρος ‘boar’ clearly point to a common protoform **ǵ^hór-jo-s-*, this is likely not a Helleno-Albanian innovation since Finn. *karjas* ‘wild boar’ suggests a loan from an otherwise unattested Proto-Germanic counterpart **garjaz* (Hyllested 2020: 412 n. 26).

⁹ It is likely that the distribution is based on the presence vs. non-presence of laryngeals, as proposed for Greek by Peters (1976): **j-* > ζ- vs. **Hj-* > -; however, other scholars see exactly the reverse distribution here (e.g. LIV²). Either way, it is significant that Greek and Albanian agree on which lexemes show which reflexes.

¹⁰ See Kortlandt 1996 for a summary of the various scholarly views regarding the Albanian material.

the meaning Lith. *jaudā* ‘seduction’) ~ Gr. *ύσμίνη* ‘battle’ < **iud^h-s-* < **iēud^h-* ‘care for, be engaged in’.

2. In both Albanian and Greek, the original clusters **tj* and **dj* underwent affrication to **ts* and **dz*, and in initial position, the former further assibilated into **s-*. In Albanian, assibilation was ultimately completed in all positions, resulting in *s* and *z*, a development which happened late enough to affect Latin loanwords. The only relevant lexemes shared by both languages involve the voiced cluster: Alb. *Zoj-z* ‘Albanian sky god’ ~ Gr. *Ζεύς* < **d̥iēus* (Mann 1952: 32) and Alb. *dhjes* ‘to shit’ (with secondary final devoicing) ~ Gr. *χέζω* ‘id.’ < **ǵ^hed-iē/o-*.
3. PIE thorn clusters with a labiovelar retain the rounding (Section 13.5.1). While this is in itself an archaism, scholars who do not believe in the Core IE thorn-cluster metathesis will see a clear shared innovation here.
4. The two languages share many developments of clusters containing sonants. For example, **-s-* was lost with compensatory lengthening before a sonant, e.g. Alb. *dorë* ‘hand’ < **ǵ^hērā* < **ǵ^hés-rā* ~ Gr. *χεῖρ* ‘id.’ < PIE **ǵ^hés-r-* and Alb. *krua* ‘spring’ m., pl. *kronj* ~ Gr. *κρήνη*, Dor. *κράνᾱ* ‘spring, well’ < **kras-neh₂* ~ PGmc. **hraznō* ‘wave’ (> OE *hærn*, ON *hrǫnn*).

13.4.7.2 Inflection and Morphosyntax

1. Under the assumption of a set of distinct past tense middle voice endings in PIE, as suggested by parallels between, e.g., Greek and Sanskrit, e.g. 3sg. *-to* ~ *-ta*, 1pl. *-μεθα* ~ *-mahī*, 3pl. *-οντο* ~ *-anta*, it is interesting that both Greek and Albanian have formations with specifically active past endings in a non-active past paradigm. That is, in the aorist passive, as opposed to middle forms with the endings given above (*-to*, etc.), Greek adds active endings to the passive stem, e.g. 1sg. *ἐπλύθην-ν* ‘I-was washed’ / 2sg. *ἐπλύθης-ς* ‘you-were washed’, etc. (for the endings, cf. active imperfect 1sg. *ἔπλυνον-ν* ‘I-was washing’ / 2sg. *ἔπλυνες-ς* ‘you-were washing’); similarly, Albanian uses active forms with the formative *u* (based on the PIE reflexive element **sue*), e.g. *u lava* ‘I-was washed’ / *u lave* ‘you-were washed’ (for the endings, cf. active past *lava* ‘I-washed’ / *lave* ‘you-washed’). These past forms with active endings are in addition, in both languages, to inherited special present medio-passive endings (e.g. 1/2/3sg. Gr. *-μαι/-σαι/-ται*, Alb. *-m/-sh/-t*). It thus appears that both have innovated to use ostensibly active endings in a past passive formation.
2. As pointed out in footnote 3, both Albanian and Greek show the inherited use of the negator **meh₁* in prohibitives. Additionally, though, both also show innovative uses of **meh₁* not found elsewhere in IE. Specifically (cf. Joseph 2013), uses of **meh₁* in negating non-finite forms (e.g. Alb. *për të mos dështuar* ‘(in order) to not fail’, Gr. *τὸ μὴ προμαθεῖν* ‘(the-state-of) not knowing beforehand’), in tentative questions (e.g. Alb. *mos e njihni?* ‘do you perhaps know him?’),

Gr. μή σοι δοκοῦμεν ‘do we perhaps seem to-you . . . ?’), and in introducing ‘fear’ complements (Alb. *kam frikë mos e kam infektuar* ‘I-have fear lest I-have infected him’, Gr. δέδοικε μή διαφθαῶ ‘he-feared lest I-be-corrupted’) are all functional innovations found exclusively in Albanian and Greek.

13.4.7.3 Verb Formation

- One of the most characteristic innovations shared by Albanian and Greek is a group of new productive verbal present types combining a nasal present and a *i*-present. They sometimes build on old nasal presents such as **h₂eub^h-n_i-* > Alb. *venj* ‘weave’, Gr. *ύφαίνω* ‘weave’ (Porzig 1954: 178; cf. Ved. *ubhnāti*), sometimes not (see Section 13.4.8 on **b^heh₂-* ‘shine’ > Alb. *běj* ‘does’, Gr. *φαίνομαι* ‘appear’). They may even be denominal, as with Alb. *thaj*, Arbëresh *thanj* ‘dry up’ ~ Gr. *αὔαινω* < **saus-n_i-*, denominative to **saus-o-* ‘dry’ (Gr. *αῶς*).
- Relatedly, both languages often create simple secondary *i*-presents for verbs with roots ending in a sonant; they share at least three such verbs:
 - PIE **ten-* ‘to stretch’: *nu*-present **t_n-néu-* (cf. Ved. *tanóti*) → **ten-je-* in Alb. *n-de(n)j* and Gr. *τείνω*
 - PIE **der-* ‘tear apart’: thematic present **der-e-* → **der-je-* in Alb. *djerr* ‘destroy’ ~ *δείρω* (alongside *δέρω*) ‘to skin, flay’ (pace Orel 1998: 69 and LIV² 119–20)
 - PIE **d^hg^{wh}er-* ‘flow; diverge, perish’: thematic present **d^hg^{wh}er-e-* → **g^{wh}per-i-* (cf. Section 13.5.1 and compare Ved. *kṣárati* ‘flow; wane, perish’, Av. *γžaraiti* ‘flow’).
- As mentioned in Chapter 12, a new type of *s*-aorist arose in the broader Balkanic subgroup already, formed with **eh₂-s-* to denominative verbs in **-eh₂-je-*. By analogy, Albanian and Greek agree on forming an *s*-aorist to the PIE root **deh₂i-* ‘share, divide’, cf. Alb. *(n-)dava*, Gr. *ἐδαισάμην* ‘I shared’ vs. the old root aorist in Ved. *(ava) adāt* ‘split off’ (LIV² 103–4).
- The OAlb. 3sg.aor. *u n-gre* ‘arose’ reflects the same innovated thematic aorist **h₁gr-e/o-* as Homeric Gr. *ἐγρετο* ‘woke up’, to the root **h₁ger-* ‘wake up’, replacing an original athematic aorist (Schumacher 2017).
- Several verbs co-occur with **peri-* ‘around’ in both languages:
 - *peri-k^ul-n-h₁-* ‘turn around’ > Alb. *për-kul* ‘to bend, curve’ ~ Gr. *περι-τέλλομαι* ‘go in circles’ (LIV² 386)¹¹
 - *peri-seh₂g-* lit. ‘drive around’, lexicalized as *për-gjoj* ‘listen closely; eavesdrop’ ~ Gr. *περι-ηγέομαι* ‘explain, describe’ (alongside ‘lead around’)

¹¹ Although the context in which OIr. *do-air-chella* ‘conceals’ is attested also allows for a translation ‘encloses (of water)’, *ar-cela* alone means ‘takes away, steals’, and it rather contains the PIE root **kel-* in *celim* ‘hides’ (Edel 2006: 83 n. 46; Le Mair 2011).

- c. **peri-pek^w*- ‘bake all over’, lexicalized as ‘crust over’ > Alb. noun *për-peq* ‘colostrum pudding’, secondary from the pl. of **për-pak* ~ Gr. *περι-πέσσω* metaph. ‘gloss over, cajole’.
6. The Albanian copula is prefixed with **h₁en-*: Geg *âsht* ~ Tosk *është* ‘is’ < **h₁en-h₁esti* corresponding to Gr. *ἔνεστι* ‘is in’ alongside short forms in Tosk *ë* and Koine *ἐν* (cf. Hamp 1980; Joseph 2016).

13.4.7.4 Nominal Formation

1. Across IE, for deriving adjectives from **sal-* ‘salt’, various suffixes are found, e.g. **-iko-* in Germanic (e.g. NHG *salz-ig*), **-no-* in Slavic (e.g. Ru. *sol-ën-yj*), but both Albanian and Greek show parallel formations with an **-m-* suffix alone or together with **-i-*: Alb. *n-gjel-m-ët* ‘salty’ ~ Gr. *ἄλιμος* ‘of the sea’, *ἀλ-μ-υρός* ‘briny’.
2. Based on the need for **ā* or **ē* in the preform of Albanian *sot* ‘today’, in order to motivate the *o*-vocalism, Joseph (2013) posits a pre-Albanian adverbial composed of a deictic element **k_i* with **āmer* for ‘day’, **k_i-āmer-*, ‘this day’; later, after a metanalysis to **k_jā-mer-*, the more usual word for ‘day’, **diti-*, replaced **(ā)mer*, giving **k_jā-diti*, from which *sot* developed regularly. This lexeme occurs also in Greek (cf. *ἦμαρ*, *ἡμέρα*) and Armenian (*awr*), so its presumed occurrence here may link Albanian, Greek, and Armenian, but the use of this form in the word for ‘today’ specifically links Albanian and Greek, since Greek has *σήμερον* (Attic *τῆμερον*) < **k_j-āmer-o-m*.¹²
3. Alb. *bot* ‘someone; person’, *botë* ‘world; humanity; others’ < a concretized acrostic *t*-stem noun **b^hueh₂-t-* ‘living being’ < abstract ‘becoming’ ~ **b^hueh₂-t-éh₂*, collective of **b^hueh₂-t-ó-* ‘having life’, respectively ~ Gr. *φός*, gen. *φωτός* ‘man; mortal’ < **b^huoh₂-t-* (Kashima 2019).
4. Alb. *huaj* ‘stranger (sb.); foreign, alien (adj.)’ formally matches Gr. *ξένιος*, an epithet of Zeus derived from *ξένος*, Dor. *ξένφος*, Ion. *ξείνος* (‘id.’; Porzig 1954: 178). The protoform **ksen_uo-* < **g^hs-en-_uo-* contains the same root as NW IE **g^hós-ti-s* ‘guest; host’. The lengthening in Albanian (*-ua-* < **-ō-* < **-ē-*) is compensatory from the loss of **-u-* (< **ksēnja-* < **ksennja-* < **ksenujo-*; Hyllested 2013).
5. A new term **g^hersos* ‘dry land, fallow land’ from the root **g^hers-* ‘stiff’ > Alb. *djerr* ~ Gr. *χερός* (curiously reminiscent of Italo-Celtic **tersos* ‘id.’ from the root **ters-* ‘dry’).

¹² It is tempting to see the metanalysis to **k_jā-* as a shared Albanian–Greek feature, since Greek shows the same development; cf. Mycenaean *za-we-te* ‘this year’, from **k_jā-wetes* (note later *σῆτες*, Attic *τῆτες*).

6. A derivative **spor-eh₂* ‘seed; semen’ from the root **sper-* ‘spread, strew’ > Alb. *farë* ~ Gr. *σπορά*.¹³
7. A result noun **ġ^hud-tlo-* from the root **ġ^heyd-* ‘pour’: Alb. *dyllë* ‘wax; sap’, Gr. *χυλός* ‘juice’ (Porzig 1954: 178; Huld 1984: 165). The lengthening reflected in Alb. *-y-* is compensatory from the loss of **-d(s)t-*, not a sign of Winter’s Law in Albanian (cf. Section 13.4.1).
8. An instrument noun **ġemt-trom* ‘stinger’ > Alb. *thundër* ‘hoof’ (with *-un-* from **-em-* as in *tundoj* ‘tempt’ ⇐ Lat. *temptō*; same root as in Alb. *thua* ‘nail’ and *thumb* ‘bee’s stinger; thorn; arrowhead point’) ~ Gr. *κέντρον* ‘point, goad, nail’ (borrowed into Geg as *çandër, qandër* ‘forked shoring pole; prop’).
9. A derivative **h₃od-meh₂* ‘smell’ > Tosk *amëz*, Geg *amë* ‘scent; flavour’ ~ Gr. *ὀδμή* ‘stench’ vs. Lat. *odor* ‘smell’, Arm. *hot* ‘smell; savour’ (Huld 1984: 165).
10. Hamp (2015: 15) found a common collocation in Alb. *bie erë* ‘smell’ < **b^her-* + **h₃od-r-eh₂* vs. Gr. *ὀσφραίνουμαι* ‘to smell’ < **h₃od-s-* + *b^her-* lit. ‘carry odour’.
11. The name of the Albanian dawn-goddess, goddess of love and protector of women, *Premtë, P(ë)rende* corresponds regularly to the Greek name *Περσέφασσα*, a variant of *Περσεφόνη*, which Janda (2000: 224–50) convincingly traces back to **pers-é-b^h(h₂)nt-ih₂* ‘she who brings the light through’. The development of *-b^hηC-* would be the same as in *venj* ‘weave’ < **vemj-* < **h₂eub^h-ñi-* (cf. Section 13.4.7.1 (1)); regarding Alb. *-r-* from originally pretonic *-rs-*, cf. *ter* ‘to dry’ from the PIE causative **tors-éje-*.
12. In both Albanian and Greek, two PIE *u*-stems, **ġén-u* ‘knee’ and **dór-u* ‘tree’, occur with *-n*-extensions: Alb. *gju* ‘knee’, Geg *gjú*, def. *gjúni* (cf. Section 13.4.3) and *dru* ‘tree’, Geg *drú*, def. *drúni* ~ Gr. *γόνατον* (alongside original *γόνη*) and *δόρ(φ)ατος* (Huld 1984: 165).
13. PIE **h₂end^hos* ‘meadow vegetation’ acquired the meaning ‘flower’ in both Alb. *endë* and Gr. *ἄνθος* vs. Arm. *and* ‘field’, Ved. *ándha-* ‘herb’, Toch.B *ānt* A *ānte* ‘plain’ (Huld 1984: 164; Kortlandt 1986: 39). From this noun, a new verb **(h₂)and^h-éje-* was derived, yielding Alb. *ëndem*, Gr. *ἀνθέω* ‘blossoms’. Formally, they correspond to Arm. *andem* ‘cultivate’ (Danka & Witczak 1995: 124), but the meaning differences suggest that the Armenian derivation happened independently.
14. The Albanian *o*-grade derivative *darkë* ‘supper, dinner; evening’ matches Gr. *δόρπον* ‘evening meal’ < **dórk^wom* (Porzig 1954: 178; Jokl 1963:

¹³ Alb. *farë* meaning ‘affinity; kind’ is historically a different word, borrowed from Langobardic *fara* ‘military clan’ into almost all Balkan languages, including Romanian, Bulgarian, and Modern Greek.

- 154); the root is not isolated if akin to Bret. *dibri, dribi* ‘eat’ (per Hamp 1966).
15. It has long been known that Alb. *për-pjetë* ‘steep’, prep./adv. ‘upwards’, noun f. ‘hill, slope’, from **pro-peth₂-o-* corresponds accurately to Gr. *προπετής* ‘falling forwards’, containing the root of *πέτομαι* ‘fly’ (Orel 1998: 321 with references). But it has gone unnoticed that the phrase underlying the counterpart *tatë-pjetë* ‘slope; (adv.) downhill’ (Orel 1998: 450) also occurs in Gr. *κατα-πίπτω* ‘fall down’.
 16. If Nikolaev (2009: 195) has correctly derived Arm. *learn* ‘mountain’ and OIr. *lie* ‘stone’ from **lēh₂u-r-*, **lēh₂u-n-*, then Albanian and Greek agree on a secondary thematic derivative **leh₂u-r-eh₂* ‘rockfall’ > Alb. *lerë* ‘boulder; stone heap’ ~ Gr. (Attic) *λαύρα*, Ep. Ion. *λαύρη* ‘narrow passage, alley’ (so too Jokl 1934: 46–8).¹⁴
 17. Albanian and Greek agree on a *-no-*derivative **k₂uap-nó-s* ‘smoke’ > Gr. *καπνός* ‘smoke’, Alb. *kem* ‘incense’ vs. other derivatives in Lat. *vapor* ‘steam’, Lith. *kvāpas* ‘breath; smell’ (Porzig 1954: 177).
 18. An *-i-* in the stem of **kōuH-(i-)lo-* ‘hollow; empty’ is reflected only in Alb. *thellë* ‘deep; dark(-coloured)’, Gr. *κοῖλος, κóϊλος*, Myc. *ko-wi-ro* ‘hollow’ (Porzig 1954: 177; differently Huld 1978).¹⁵
 19. PIE **g^welH-* ‘torment, sting’ in words for ‘sewing needle’ > Alb. *glep, gjep, gjilpërë*, Geg *gjylpânë* ~ Gr. *βελόνη* (Irslinger 2017: 312). The Albanian suffix *-ërë, -ânë* even formally matches Gr. *-ónē* < **-mn-eh₂* (Olsen 1999: 492; Rasmussen 1996: 154), used in denotations for instruments and remedies.
 20. Alb. *bar n.*, pl. *barëra*, Geg *barna* ‘grass; medicine’ ~ Gr. *φάρμακον* ‘drug, medicine’ < **b^har-(m)n-* (Jokl 1963: 129), derived from the Core IE root **b^har-* which denotes crops everywhere else (e.g. Lat. *far* ‘spelt’, Eng. *barley*).
 21. Alb. *ndër-dym* ‘in doubt’ formally corresponds to Gr. *διά* ‘apart, through’ < **d₂uis-m̥* ‘in two (parts)’ (Mann 1952: 32).
 22. A pronoun **h₂a₂uto-* ‘self’ occurs in Alb. *vetë*, Gr. *αὐτός* (Witzczak 1997: 216); also shared with Phrygian (*avtos*; see Section 11.4.2).

13.4.7.5 Semantic Innovations (Selection)

1. PIE **seh₂g-* ‘seek’ → ‘drive’: Alb. *gjuaj* ‘drive (quickly), chase’, Gr. *ηγέομαι* ‘lead the way, guide’ (cf. Section 13.4.7.3 (5b)).

¹⁴ Milyan *lakre* is formally identical to the Helleno-Albanian word, but possibly means ‘stone tablet’ (Nikolaev 2009: 196).

¹⁵ Arm. *soyl* ‘cave’ appears to be a ghost form and would reflect **kōuH-lo-* anyway (Zair 2011: 166 n. 5).

2. **lógh-o-* ‘resting-place’ (Slavic **logъ* ‘lair’, Toch. B *leke* ‘bed’) → ‘camp’ → ‘troop, band’ in Alb. dial. *lag*, Gr. *λόχος* (Hyllested 2020: 410–11).
3. *b^húh₂-m̃* ‘growth’ (Ved. *bhūman* ‘world, region (n.); multitude, wealth’ (m.)) → ‘plant’ in Alb. *bimë*, Gr. *φύμα* (Mann 1950: 387).
4. **h₂end^hos* ‘meadow vegetation’ → ‘flower’ (Section 13.4.7.4 (13)).¹⁶
5. **h₁erǵ^h-* ‘go; jump up’ → ‘come’ in Alb. *erdh-* aor., Gr. *έρχομαι*.
6. **h₁éh₁tr̥* ‘stomach; intestines’ (PGmc. **ēbrō* ‘veins, entrails’, e.g. > OE *ædre*, also ‘sinew; kidney’) → ‘heart’: in Alb. *votër, vatër*¹⁷, Gr. *ήτορ*.
7. **kras-neh₂* ‘wave’ (Section 13.4.7.1 (4)) → ‘spring, well’ in Alb. *krua*, Gr. *κρήνη*; compare Eng. *well* ~ NHG *Welle* ‘wave’, Lith. *vilnis*, Ru. *volná* ‘id.’

13.4.8 A Palaeo-Balkan Group?

Evidence for a broader Balkanic group consisting of Albanian, Greek, and Armenian, as well as Phrygian, is presented in Section 12.4.1 and (mainly) Section 12.5.¹⁸ To this we can add

1. A new possessive pronoun **emos* ‘mine’ > Alb. *im(e)*, Gr. *ἐμός*, Arm. *im*, perhaps dissimilated from an old accusative *me-me* (Huld 1984: 165 with references).
2. A suppletive aorist **g^werh₃-* to the verb ‘eat’, irrespective of the origin of the present stem. Compare Alb. *ha*, aor. *n-grë*; Gr. *έδω, έσθίω*, aor. *έφαγον, έ-βρώ-θην*; Arm. *owt’em*, aor. *k’er-* (Holst 2009: 87).
3. By the same analogy described in Section 13.4.7 Verb Formation (3), the old root aorist of PIE **steh₂-* ‘stand’ was replaced with an *s*-aorist **steh₂-s-* with factitive semantics in both Alb. *shtova* ‘added’, Gr. *έστησα* ‘made stand’, Arm. *stac’ay* ‘acquired’, Phryg. *estaes* ‘erected’, and Mess. *stahan* ‘erected’ (Søborg 2020: 76).
4. A new root **klaui-* ‘to cry’ > Alb. *qaj*, OAlb. *klanj* < **klaui-ni-* ~ Gr. *κλαίω*, Arm. *lam* ‘to cry’.
5. The originally honorific term **h₂ner-* ‘man (of consequence)’ has replaced **uiHró-* as the common word for ‘man’, Alb. *njeri*, Gr. *άνήρ*, Arm. *ayr* (Huld 1984: 165).
6. Generalized full-grade in the word for ‘louse egg’: Alb. *thëri*, Geg *thëni* < **konid-*, Gr. *κονίς* and Arm. *anic* (dissimilated from **sanic*) vs. zero-grade

¹⁶ Changes in specific plant-names (e.g. Alb. *ah* ‘beech’ ~ Gr. *όζύα* ‘id.’ vs. ‘ash tree’ elsewhere) are not included here as they may reflect new geographical surroundings rather than genealogy.

¹⁷ Synchronically identical to *votër, vatër* ‘fireplace, hearth’ (understood as the middle of the house) due to merger with PIE **h₂eh₁-tr̥* ‘id.’.

¹⁸ We can embrace most of the evidence adduced there although we note that (1) Alb. *edh* ‘goat’ may simply be borrowed from Lat. *haedus*, cf. Rom. *ied* (Witczak 1997: 125); (2) the locative plural ending *-*si* is not secured for Albanian since even *-*su* may yield the attested outcome -*sh*; and (3) *awr* ‘day’ etc. was probably not originally restricted to Greek and Armenian (Section 13.4.7.4 (2)). On Alb. *grua* ‘woman’, see also Opfermann (2017).

**knidā* in Germanic and Balto-Slavic: OE *hnitu*, Latv. *gnīda*, Scr. *gnjīda* (Huld 1984: 165).

7. **ster-ih₂* ‘sterile (of females)’ > Alb. *shtjerrë*, Gr. *στεῖρα*, Arm. *sterj* (Hyllested 2016; on the Greek-Armenian connection see Lamberterie 2013).

8. Perhaps PIE **k^{we}ǵ-* ‘gather’ > Palaeo-Balkan ‘make’ (Section 13.4.2).

There is also some evidence for a broader Balkanic unity wherein further developments set Albanian and Greek apart from Armenian, again pointing to a Helleno-Albanian subgroup:

9. PIE **b^heh₂-* ‘shine’ (LIV² 68–9) forms a nasal present in Albanian, Greek and Armenian, but only Albanian and Greek add an extra *i*-present to it, following a productive pattern (Section 13.4.7.3 (1)): Armenian *banam* < **b^heh₂-n-* vs. Alb. *běj*, Geg *bāj* ‘does’, Greek *φαίνομαι* ‘appear’ < **b^heh₂-n-ǵ-*.

10. A derivative **Hon-ǵ-jo-* (alongside archaic **Hon-ǵ-*) ‘dream’ occurs in Alb. *ëndërr* ~ *ëndër* and Gr. *ὄνειρος* ~ *ὄναρ* vs. Arm. *anowrj* (< **Hnōr-ǵo-*), all ‘dream’ (Lamberterie 2013; Kortlandt 1986: 38; Witczak 1997: 126). Its root is not found elsewhere, but the heteroclitlic declension points to an IE retention in Palaeo-Balkan.

11. A derivative **h₁ed-ún-eh₂* ‘pain’ > Alb. *dhunë*, *dhurë* f.pl. ‘damage, injury; shame, disgrace’ = Gr. *ὀδύνη* ‘pain’ alongside the older **h₁ed-ũn-* in Arm. *erkn* ‘labour pains’ and e.g. secondary **h₁ed-ōn* in OIr. *idu* (not **-ũn-* since **-du-* > OIr. *-db-*).

And in one case, an Armenian innovation isolates it from a Helleno-Albanian remainder:

12. The word for ‘bee’ is derived from **mél-it* ‘honey’ in all three languages (Holst 2009: 90): Alb. *mjaltë* ‘honey’ ~ *bletë*, *mjalcë* ‘bee’, Gr. *μέλι* ‘honey’ ~ *μέλισσα*, *μέλιττα* ‘bee’, Arm. *mełr*, *-ow* ‘bee’ ~ *melow* ‘honey’, but Armenian has *-u-* by influence from PIE **méd^hu* ‘mead’ (Clackson 2017: 112).

13.5 The Position of Albanian

13.5.1 Broader Connections within IE: Albanian and the Centum–Satem Division

Starting with reconstructed PIE with a three-way distinction in the guttural consonants (palatals, e.g. **ǵ*, velars, e.g. **k*, and labiovelars, e.g. **k^w*), a division within IE is possible, descriptively, into branches that merge palatals and velars (so-called *centum* languages) and those merging velars and labiovelars (*satem* languages). The *satem* languages also show affrication and/or assibilation of the PIE palatals. We say “descriptively” because we do not see this division as a basically genealogical one within IE. For us, the *centum* languages are not a coherent dialectal or genealogical subgroup though the *satem* languages might be. The position of Albanian within this

scheme is thus of considerable interest and, not surprisingly, is somewhat complicated.

In particular, while Albanian shows some merger of labiovelar and velar, e.g. *pjek* ‘to cook’ < **pek*^w- (cf. Gr. *πέπων* ‘ripe’) and *plak* ‘old man’ < **plak*- (cf. Lith. *pilkas* ‘grey’), it also maintains the original three-way guttural distinction in some environments, and thus descriptively is neither *centum* nor *satem*. As recognized by Pedersen 1900, they all show distinct outcomes before original front vowels, e.g. *tho-të* ‘says’ < **kē*-*ti* < **keh*₁-*ti* (cf. Old Persian *ṣā-tiy*), *kohë* ‘time’ < **kēs*ko- (cf. OCS *časъ* ‘hour’), and *sorrë* ‘crow’ < **k^wersno*- (a vṛddhi derivative of ‘black’, cf. Sanskrit *kr̥ṣṇá-*). In this way, Albanian behaves like Luvian, as analyzed by Melchert 1987. Moreover, since elsewhere in Anatolian, *centum*-like mergers happened independently (e.g. Hitt. *kī-ta* ‘lies’ < **kē*₁-, cf. Ved. Br.+ *śé-te*), *centum*-ness cannot be considered a significant innovation. In fact, *centum*-ness seems relevant only for post-Anatolian and post-Tocharian IE, and really equates to just Italo-Celtic and Germanic; *satem*-ness, by contrast, equates to Balto-Slavic and Indo-Iranian (and could be a real shared innovation between them). An ancient Balkan group, including Armenian, Albanian, and Greek, appears like a potpourri, making up a third unit which initially kept all original stop distinctions; various developments in its individual sub-branches subsequently obscured this basic retention, e.g. the Albanian **k*/**k^w* merger in some environments noted above, or the assibilation seen in *sjell* ‘bring’ < **k^wel*- (cf. Gr. *πέλω* ‘be in motion’).

Albanian lexemes with initial clusters *vd-* and *ft-* are of special interest in this context. Previous etymologies of the two clearest examples, Alb. *vdjerr* ‘to disappear’ and *vdes* ‘to die’ (stem *vdek-* as in the participle *vdekur* ‘dead’), all involve a semantically vague labial prefix *v-* supposedly added to known verbal stems (e.g. Mann 1952; Orel 1998; Hamp 2004; Holm 2011). However, a less dichotomous *centum*–*satem* division, with Balkan languages showing characteristics of both, allows for a more economical analysis of these Albanian verbs as regular reflexes of Core IE “thorn clusters” containing a labiovelar. Thus, Alb. *vdjerr* can simply correspond fully to Gr. *φθείρω* ‘destroy, ruin’, med. *φθείρομαι* ‘perish’, even down to the *i*-present, from Core IE **g^{wh}per*- < PIE **d^hg^{wh}er*- ‘flow; melt away; disappear’, and no prefix need be posited. Similarly, *vdes* could straightforwardly contain the Core IE root **g^{wh}pej*- < PIE **d^hg^{wh}ei*- ‘decline; perish’ also seen in Gr. *φθί(ν)ω* ‘perish’, *φθίμενοι* ‘the dead’, though formally from a causative **g^{wh}poi-k^w-éje*- ‘leave behind’ (→ ‘depart’).¹⁹

An important consequence of this interpretation is that, since Albanian *v-* or *f-* reflects the old labiovelar, the dental *-d-* must continue the PIE thorn element. This, in turn, would mean that the common view that Albanian agrees with

¹⁹ A candidate for a reflex of the unvoiced counterpart **tk^w*- might be Alb. *ftik* ‘dry’ ~ Lat. *siccus* ‘dry’ (< **siccus*), if from a PIE **tk^wiH-ko-* or **tk^wei-ko-*, possibly also reflected in PGmc. **swīþan*- ‘scorch’ and/or Gr. *ψι-λός* ‘bare’. None of these words have generally accepted etymologies.

Balto-Slavic, Germanic, and Italic in preserving only the dorsal part of *palatal* thorn clusters – as if **ǵ^hpōm* ‘earth’ and **ǵ^hpjes* ‘yesterday’ were **ǵ^hōm* and **ǵ^hies*, respectively – must be abandoned. Although the regular reflex of a palatal **ǵ^(h)* in Albanian is *d(h)*- as well, the sole consonant left in *dhe* ‘earth’ and *dje* ‘yesterday’ must then reflect the thorn element and not the dorsal, which disappears without a trace.

The above analysis has important consequences for the internal classification of IE:

1. It makes Albanian more of a *centum* language, since it preserves not only the velar-labiovelar distinction but even the actual rounding of labiovelars.
2. It distances Albanian from Balto-Slavic, Germanic, and Italic, which all agree on preserving the stop part of thorn clusters only.
3. It connects Albanian even more to Greek than previously assumed.

13.5.2 Conclusion

As noted at the outset, the relationships that Albanian shows within IE are complicated, and the evidence discussed here should make that point abundantly clear. We have surveyed the most striking possible connections that Albanian shows with other branches of Indo-European, based on key pieces of evidence.²⁰ Technically speaking, from a genealogical standpoint, Messapic likely is the closest IE language to Albanian (Matzinger 2005). However, in the absence of sufficient evidence, that connection must remain speculative. Among the other connections, leaving aside the broad *centum*–*satem* parameter, since we do not see it as a valid dialect division in the usual sense, we are left with the following, listed from the least compelling (with Italic) to the most compelling (with Greek):

- Albanian and Italic
- Albanian and Celtic
- Albanian and Indo-Iranian
- Albanian and Germanic
- Albanian and Balto-Slavic
- Albanian and Armenian
- Albanian and Armenian, Greek, Phrygian, and Messapic (etc.)
- Albanian and Greek

These are not necessarily mutually exclusive, depending on one’s overall conception of the interrelationships among all branches of IE. That is, some apparent

²⁰ We have deliberately restricted ourselves to the best evidence, leaving out some intriguing shared substratum words such as Alb. *dëllinje*, *dëlli* ‘juniper’ ~ Gr. (Hsch.) *σχέλινος* ‘wild cypress or juniper’, indicating a protoform **(s)ǵ^helin-(i)o-* (Danka & Witczak 1995: 132); and formations containing isolated roots such as **uisǵ^h-i(i)o-* > Alb. *vithe* ‘haunch, especially of a horse’ ~ Gr. *ισχίον* ‘hip-joint; loins, haunch’ (Mann 1952: 39).

shared innovations could in principle result from wave-like diffusion in prehistoric times. Moreover, as noted throughout, one has to ask whether limited evidence for a particular linkage goes beyond what any two branches might show.

Ultimately, though, as indicated, the preponderance of evidence favours a close connection between Albanian and Greek,²¹ possibly as a subset within a “Palaeo-Balkanic” group with Armenian and Greek, as well as Phrygian, Messapic, and other fragmentarily attested languages (see Figure 13.1). The Albanian–Greek connection that we argue for here is particularly interesting in the light of the computational phylogenetic study of the interrelationships among IE languages reported on in Chang et al. 2015. In that paper, starting with the same model and data set as earlier phylogenetic studies (especially Bouckaert et al. 2012, 2013), but with a key difference in that they “constrained eight ancient and medieval languages to be ancestral to thirty-nine modern descendants” to allow for greater accuracy, the authors develop an “analysis with modern languages from all IE subfamilies” (Chang et al. 2015: 199–200) in which Albanian, represented by Arvanitika and Tosk,²² ends up in their resulting tree diagram of IE relationships as being most closely connected to Greek. Different methods and different IE data sets and different assumptions can of course yield different results,²³ but we take heart from the convergence of our more traditional qualitative assessment of Albanian’s closest relative and the computational quantitative assessment by Chang et al.

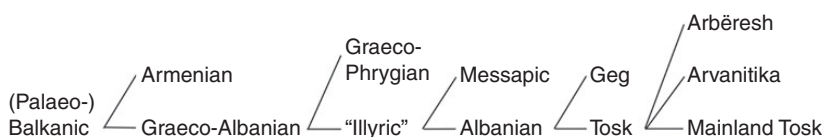


Figure 13.1 The position of Albanian²⁴

²¹ In line with our interest in just presenting the best evidence, we have focused on shared innovations. However, shared retentions can in principle, if unusual enough compared to the rest of the family, and especially when paired with significant shared innovations, point to a close genealogical connection; in a certain sense, retaining something can, under appropriate circumstances, be innovative in itself. See also footnote 5.

²² Arvanitika, of course, is a Tosk dialect, but we assume that by “Tosk”, Chang et al. mean the standard language, which is based on a Tosk variety.

²³ Ringe, Warnow, and Taylor (2002), for instance, as noted in Section 13.4.5, see Albanian and Germanic as particularly closely related.

²⁴ In the absence of linguistic data about ancient Illyrian, we feel caution is in order about the connection between Illyrian, whatever that label might have meant to the ancients, and Albanian, even if that connection might be reasonable from a geographic and archaeological perspective (so Katičić 1976).

References

- Bouckaert, Remco, et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337. 957–60.
- Bouckaert, Remco, et al. 2013. Corrections and clarifications. *Science* 342. 1446.
- Çabej, Eqrem. 1969. Keltisch-albanische Isoglossen. In *Studi linguistici in onore di Vittore Pisani*. Vol. 1, 167–86. Brescia: Paideia.
- Çabej, Eqrem. 1975. Zu einigen slawisch-albanischen Isoglossen. In F. V. Mareš, Bonifacy Miazek & Katja Sturm-Schnabl (eds.), *Bereiche der Slawistik: Festschrift Josip Hamm*, 37–51. Vienna: VÖAW.
- Chang, Will, Chundra Cathcart, David Hall, & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91. 194–244.
- Clackson, James. 2017. Contamination and blending in Armenian etymology. In Bjarne Simmelkjær Sandgaard Hansen et al., *Usque ad radices: Indo-European studies in honour of Birgit Anette Olsen*, 99–115. Copenhagen: Museum Tusculanum.
- Cowgill, Warren. 1960. Greek *ou* and Armenian *očʻ*. *Language* 36. 347–50.
- Danka, Ignacy Ryszard, & Krzysztof Tomasz Witczak. 1995. Some remarks on the Albanian vocabulary in the Palaeo-Balkan and Indo-European view. *Analecta Indoeuropaea Cracoviensia Ioannis Safarewicz memoriae dicata*, 123–35. Cracow: Universitas Jagellonica.
- Edel, Doris. 2006. “Bodily matters” in early Irish narrative literature. *Zeitschrift für celtische Philologie* 55. 69–107.
- Fellner, Hannes A. 2022. Polar life in the Tarim Basin. In Hannes A. Fellner, Melanie Malzahn & Michaël Peyrot (eds.), *lyuke wmer ra: Indo-European studies in honor of Georges-Jean Pinault*, 125–37. Ann Arbor: Beech Stave.
- Garnier, Romain. 2014. Italicque commun **né=xe=ajud* ‘jamais de la vie’. *Wékʷos* 1. 95–110.
- Hackstein, Olav. 2020. The system of negation in Albanian: Synchronic constraints and diachronic explanations. In Bardhyl Demiraj (ed.), *Altalbanische Schriftkultur – aus der Perspektive der historischen Lexikographie und der Philologie der Gegenwart*, 13–32. Wiesbaden: Harrassowitz.
- Hamp, Eric P. 1966. The position of Albanian. In Henrik Birnbaum & Jaan Puhvel (eds.), *Ancient Indo-European dialects*, 97–121. Berkeley: University of California Press.
- Hamp, Eric P. 1980. Albanian *ështëë*. In Kathryn Klar, Margaret Langdon, & Shirley Silver (eds.), *American Indian and Indoeuropean studies: Papers in honor of Madison S. Beeler*, 337–46. The Hague: Mouton.
- Hamp, Eric P. 1982. The oldest Albanian syntagma. *Balkansko ezikoznanie* 25. 77–9.
- Hamp, Eric P. 1992. Albanian. In Jadranka Gvozdanović (ed.), *Indo-European numerals*, 835–921. Berlin: Mouton de Gruyter.
- Hamp, Eric P. 2004. Indo-European **su-* in Albanian. In Adam Hyllested et al. (eds.), *Per aspera ad asteriscos: Studia Indogermanica in honorem Jens Elmegård Rasmussen sexagenarii Idibus Martii anno MMIV*, 207–9. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Hamp, Eric P. 2015. *Albanian historic syllabics*. Oxford: University of Mississippi.
- Holm, Hans J. 2011. “Swadesh lists” of Albanian revisited and consequences for its position in the Indo-European languages. *Journal of Indo-European Studies* 39. 1–53.

- Holst, Jan Henrik. 2009. *Armenische Studien*. Wiesbaden: Harrassowitz.
- Huld, Martin. 1978. Albanian *i thellë* and its eastern relations. *Zeitschrift für vergleichende Sprachforschung* 92. 297–9.
- Huld, Martin. 1984. *Basic Albanian etymologies*. Columbus, OH: Slavica.
- Hyllested, Adam. 2012. Albanian *hundë* ‘nose’, and Faroese, SW Norwegian *skon* ‘snout’. In Stephanie W. Jamison, H. Craig Melchert, & Brent Vine (eds.), *Proceedings of the 23rd annual UCLA Indo-European conference*, 73–81. Bremen: Hempen.
- Hyllested, Adam. 2013. Unexpected lengthened grade in Albanian. Paper presented at the Lengthened Grade in Indo-European conference, Leiden University, 29–31 July 2013.
- Hyllested, Adam. 2016. Proto-Indo-European reconstruction and Albanian phonotactics. In Stephanie W. Jamison, H. Craig Melchert, & Brent Vine (eds.), *Proceedings of the 26th annual UCLA Indo-European conference*, 63–81. Bremen: Hempen.
- Hyllested, Adam. 2020. First-degree kinship terms in Albanian: From prehistory to the present. In Benedicte Nielsen Whitehead, Birgit Anette Olsen, & Janus Bahs Jacquet (eds.), *Kin, clan and community in Indo-European society*, 397–422. Copenhagen: Museum Tusculanum.
- Irslinger, Britta. 2017. The ‘sewing needle’ in Western Europe: Archaeological and linguistic data. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Usque ad radices: Indo-European studies in honour of Birgit Anette Olsen*, 307–23. Copenhagen: Museum Tusculanum.
- Janda, Michael. 2000. *Eleusis: Das indogermanische Erbe der Mysterien*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Jokl, Norbert. 1927. Kelten und Albaner: Eine Untersuchung ihrer Isoglossen. In *Symbolae grammaticae in honorem Ioannis Rozwadowski*. Vol. 1, 235–50. Cracow: Jagiellonian University.
- Jokl, Norbert. 1934. Zur Lehre von den alb.-griech. Teilgleichungen. *Revue internationale des études balkaniques* 1. 46–64.
- Jokl, Norbert. 1963. Die Verwandtschaftsverhältnisse des Albanischen zu den übrigen indogermanischen Sprachen. *Die Sprache* 9. 113–56.
- Joseph, Brian D. 2005. Some ancient shared metaphors in the Balkans. *Studia Albanica* 42. 45–8.
- Joseph, Brian D. 2013. On old and new connections between Greek and Albanian: Some grammatical evidence. In Aristotel Spiro (ed.), *Papers from the first international conference of Greek-Albanian studies*, 7–21. Tirana: Albanian-Greek Association of Philology.
- Joseph, Brian D. 2016. Balkan, Indo-European, and universal perspectives on ‘be’ in Albanian. In Andrew Miles Byrd, Jessica DeLisi, & Mark Wenthe (eds.), *Tavet tat satyam: Studies in honor of Jared S. Klein on the occasion of his seventieth birthday*, 130–7. Ann Arbor, MI: Beech Stave.
- Joseph, Brian D. 2022. Here’s to a long life! Albanian reflections of Proto-Indo-European semantics. In Michele Bianconi & Marta Capano (eds.), *Contact, variation, and reconstruction in the Ancient Indo-European languages: Between linguistics and philology*, 301–12. Leiden: Brill.
- Kashima, Keita. 2019. Nominal evidence for the PIE verbal root **b^hueh₂-* ‘become’. In David M. Goldstein, Stephanie W. Jamison, & Brent Vine (eds.), *Proceedings of the 30th annual UCLA Indo-European conference*, 123–39. Bremen: Hempen.

- Katičić, Radoslav. 1976. *Ancient languages of the Balkans*. 2 vols. The Hague: Mouton.
- Klingenschmitt, Gert. N.d. *The Albanian numerals*. Unpublished manuscript.
- Kortlandt, Frederik. 1986. Armenian and Albanian. In Maurice Leroy & Francine Mawet (eds.), *La place de l'arménien dans les langues indo-européennes*, 38–47. Louvain: Peeters.
- Kortlandt, Frederik. 1996. PIE **j* in Albanian. In Adrian A. Barentsen, R. Sprenger, & M. G. M. Tieleman (eds.), *South Slavic and Balkan linguistics*, 173–6. Amsterdam: Rodopi.
- Lamberterie, Charles de. 2013. Grec, phrygien, arménien: des anciens aux modernes. *Journal des Savants* 2013(1). 3–69.
- Le Mair, Esther. 2011. *Secondary verbs in Old Irish: A comparative-historical study of patterns of verbal derivation in the Old Irish Glosses*. PhD thesis, National University of Ireland, Galway.
- LIV² = Helmut Rix et al. (eds.). 2001. *Lexikon der Indogermanischen Verben*. 2nd ed. Wiesbaden: Reichert.
- Mann, Stuart E. 1950. The Indo-European vowels in Albanian. *Language* 26. 379–88.
- Mann, Stuart E. 1952. Indo-European consonants in Albanian. *Language* 28. 31–40.
- Matzinger, Joachim. 2005. Messapisch und Albanisch. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 2. 29–54.
- Matzinger, Joachim. 2016. *Die sekundären nominalen Wortbildungsmuster im Altalbanischen bei Gjon Buzuku: ein Beitrag zur albanischen Lexikographie*. Wiesbaden: Harrassowitz.
- Melchert, H. Craig. 1987. PIE velars in Luvian. In Calvert Watkins (ed.), *Studies in memory of Warren Cowgill (1929–1985)*, 182–204. Berlin: de Gruyter.
- Nikolaev, Alexander. 2009. Time to gather stones together: Greek *λάας* and its Indo-European background. In Stephanie W. Jamison, H. Craig Melchert, & Brent Vine (eds.), *Proceedings of the 21st annual UCLA Indo-European conference*, 189–206. Bremen: Hempen.
- Olsen, Birgit Anette. 1999. *The noun in biblical Armenian*. Berlin: Mouton de Gruyter.
- Opfermann, Andreas. 2017. Griechische, Armenische und Albanische 'Frauen' und die Wurzel **g^hen-*. In Harald Bichlmeier & Andreas Opfermann (eds.), *Das Menschenbild bei den Indogermanen*, 55–80. Hamburg: Baar.
- Orel, Vladimir. 1994. Albanian and Indo-European. In Jens Elmegård Rasmussen (ed.), *In honorem Holger Pedersen*, 349–64. Wiesbaden: Reichert.
- Orel, Vladimir. 1998. *Albanian etymological dictionary*. Leiden: Brill.
- Orel, Vladimir. 2000. *A concise historical grammar of the Albanian language*. Leiden: Brill.
- Pedersen, Holger. 1900. Die Gutturale im Albanesischen. *Zeitschrift für vergleichende Sprachforschung* 36. 277–340.
- Peters, Martin. 1976. Attisch *hîēmi*. *Die Sprache* 22. 157–61.
- Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. Heidelberg: Winter.
- Rasmussen, Jens Elmegård. 1985 [1987]. Miscellaneous morphological problems in Indo-European languages (I–II). *Lingua Posnaniensis* 28. 27–62.
- Rasmussen, Jens Elmegård. 1996. The origin of the Latin gerund and gerundive. *Copenhagen Working Papers in Linguistics* 4. 159–69.

- Ringe, Don, Tandy Warnow, & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Sandfeld, Kristian. 1930. *Linguistique balkanique. Problèmes et résultats*. Paris: Champion.
- Schumacher, Stefan. 2017. Old Albanian /u ngre/ ‘he/she arose’. In Bjarne Simmelkjær Sandgaard Hansen et al. (eds.), *Usque ad radices: Indo-European studies in honour of Birgit Anette Olsen*, 687–93. Copenhagen: Museum Tusculanum.
- Søborg, Tobias Mosbæk. 2020. *Sigmatic verbal formations in Anatolian and Indo-European: A cladistic study*. PhD thesis, University of Copenhagen.
- Thorsø, Rasmus. 2019. Two Balkan Indo-European loanwords. In Matilde Serangeli & Thomas Olander (eds.), *Dispersals and diversification: Linguistic and archaeological perspectives on the early stages of Indo-European*, 251–62. Leiden: Brill.
- Thumb, Albert. 1912. *Handbook of the Modern Greek vernacular: Grammar, texts, glossary*. Edinburgh: Clark.
- Winter, Werner. 1978. The distribution of short and long vowels in stems of the type Lith. *ėsti* : *vėsti* : *mėsti* and OCS *jasti* : *vesti* : *mesti* in Baltic and Slavic languages. In Jack Fisiak (ed.), *Recent developments in historical phonology*, 431–46. The Hague: Mouton.
- Witczak, Krzysztof Tomasz. 1997. Greek and Albanian: New evidence for lexical connections. *Zeitschrift für Balkanologie* 33. 215–20.
- Zair, Nicholas. 2011. Varia I: OIr *cuae*, MW *keu*, MB *queu* ‘hollow’. *Ériu* 61. 165–8.

14 Indo-Iranian

Martin Joachim Kümmel

14.1 Introduction

Indo-Iranian is mainly divided into the two big sub-branches of Indo-Aryan and Iranian.¹ Iirn. languages are first attested in the fifteenth century BCE in the Hurrian state of Mit(t)an(n)i and surrounding areas through divine, throne and personal names as well as through hippological terms. Linguistically and culturally, this variety seems to belong rather to Indo-Aryan = WIA (cf. [Mayrhofer 1982](#); [Lipp 2009](#), 1: 265–73, 310–17). Otherwise, Indo-Aryan is confined to south-eastern Afghanistan and the Indian subcontinent = (E)IA, with the language of its oldest texts, i.e. the *Rigveda*, being slightly less archaic than WIA. To explain this distribution, we can assume that IA was originally a southern branch whose speakers then migrated both westwards and eastwards, possibly under pressure from Iranian coming from the north. Iranian itself was very widespread from the Pontic steppe towards Central Asia and Mesopotamia. Its oldest texts are roughly contemporaneous with Vedic IA. Due to this wide geographical distribution, Iranian is more diverse (the validity of the sub-branch was even doubted by [Tremblay 2005](#)).

14.2 Evidence for the Indo-Iranian Branch

Of course, innovations are more interesting than archaisms. There are some important laws and changes that are characteristic for Indo-Iranian and which could be innovations.

14.2.1 Bartholomae's Law

Bartholomae's Law, i.e. the rule on progressive rather than regressive assimilation in obstruent clusters starting with a “media aspirata” stop, is an active rule in Sanskrit, and its results are still faithfully reflected in Old Avestan morphophonology, although the distinction between media and aspirata has

¹ Conventions of transcription: (P)IE **h*, **χ*, **h̥*; **k*, **g*; **q*, **g* = traditional **h*₁, **h*₂, **h*₃; **k̂*, **ĝ*; **k*, **g*.

already been lost. Accordingly, it must be reconstructed for Proto-Indo-Iranian and even Proto-Iranian.² Since there are hardly any traces of this law outside of Indo-Iranian, it is disputed whether it can be a PIE law or an IIn. innovation. However, the rule was abandoned completely as early as Younger Avestan (only isolated examples survived in later Iranian), and so the lack of evidence in IE languages attested later than that is hardly significant, since it is highly likely that the rule was lost independently. Even its absence from Anatolian and Greek may reflect rule loss, since devoicing of the aspirates in the latter would have obscured the rule, and in the former, media and aspirata merged in the same way as in Iranian and, probably, voicing was lost altogether. Furthermore, the rule is even easier to motivate in a stage of PIE that had not yet developed aspiration (cf. Miller 1977a; 1977b) and in which the “mediae” did not participate in the voicing (or fortis-lenis) contrast. Thus, it is quite possible that BL is an archaism, but its loss elsewhere is trivial enough not to require a common innovation of the other branches.

14.2.2 Grassmann’s Law

Due to the general loss of breathy voice in Iranian and Nuristanic, it is difficult to say whether Grassmann’s Law (GL), i.e. the dissimilation of an aspirated stop preceding another aspirated stop, occurred already in Proto-Indo-Iranian or only later in Indo-Aryan. The latter assumption would imply a rather long period without dissimilation, which seems quite possible considering the parallel development in Greek, where it clearly happened only after the (rather late) devoicing of aspirates.

Scharfe (1996) has argued for dialectal differences in the chronology of the application of GL and the Vedic devoicing of sibilant clusters, which would necessarily imply a late date for GL. However, this is based on very little evidence and does not explain the whole distribution (see Kobayashi 2004: 106–7, 114–16, 122–7; Lipp 2009, 1: 252–7), so it remains much more probable that GL preceded the devoicing everywhere and thus could be of PIIn. date.

There is a small circumstantial argument for an early date: the 2sg. imperative of PIIn. **ǵʰan/gʰn-* ‘to hit, kill’ starts with a palatal in both Vedic *jahí* and Avestan *jaʰdi*, while the parallel imperative of **gam-/gm-* ‘to come’ is Vedic *gahí* = Av. *gaʰdī* with the expected velar. The palatal in the former might have been taken over from the strong stem to avoid homonymy of these forms. If this had happened already in PIIn., it would presuppose that the two forms **ǵʰadʰi* ‘hit!’ and **gadʰi* ‘come!’ had already become homonymous by GL, so that

² Tremblay (2005) even took this as an argument that the original Old Avestan language still had two distinctive voiced stop series, i.e. preserved voiced aspirates.

**gad^{hi}* was replaced by **ǵad^{hi}* to solve this problem. However, a parallel development is not completely excluded: a partial spread of the palatal can also be observed in other zero-grade forms of **ǵan-*, too, cf. Ved. prs.2pl. *hathá*, OAv. infinitive *jaⁱdiiāi*.

Furthermore, there is evidence in Tocharian that it also underwent the same kind of dissimilation (but see the more cautious assessment in Section 6.5.2 n. 10): while **d^{hi}* normally became *t* (> *c* when palatalized) and thus merged with original **t*, it sometimes shows the result *ts* (palatalized), merging with **d*, and such cases only appear if a second aspirate follows, e.g. Toch.B gerundive *tsikale* < ‘should be made’ < PToch. **tsik-a-* < PIE **d^{hi}igh-*, to **d^{he}ig^h-* ‘form’. For the other stops, the eventual complete merger of all series makes it impossible to see if there was a similar dissimilation.

As a sporadic or narrowly conditioned change, aspiration dissimilation is also found in Latin (see Weiss 2018 and Section 8.2 n. 11) and Armenian (only before a nasal cluster? Cf. Rasmussen 1989: 170–1 n. 16; Martirosyan 2010: 726). In later Indo-Aryan, similar dissimilations also happened again, when new sequences of breathy voiced stops had arisen.

14.2.3 Brugmann’s Law

Brugmann (1876) postulated a change of “**a₂*”, i.e., **o* > (**ō* >) **ā* in open syllables before a consonant. This proposal did not gain much support subsequently, and Brugmann himself withdrew it. However, the reconstruction of laryngeals led to its resurrection, since it could explain many apparent exceptions as conditioned by a lost laryngeal (see Kuryłowicz 1927: 206–7; Lubotsky 2018: 1877). Pace Kiparsky (2010: §2.3), the data are still easier to explain by applying a real sound law than by invoking a special grammatically conditioned development of “floating” **o*. The counterexamples given by Kiparsky are either invalid (because they can have original **e* or a cluster **CH*) or can be explained by inner-paradigmatic analogy (as **pári-*, **āwi-*), while **ā* in the first dual and plural of the thematic inflection is not explained by Kiparsky’s account.³

A similar change can be observed in Anatolian: accented **ó* was apparently lengthened > **ō* > Hitt. Luw. *ā* (vs. **á* > *ā*), even in closed syllables, cf. Hitt. *kānki* ‘hangs’ < **kónkej*, Luwian *hās* ‘bone’ < **χóst*.⁴ Unfortunately, it remains

³ Pyysalo (2013: 114–25) rejects the law in its original form but assumes a “corrected” version, “Brugmann’s Law II”, where lengthening is only found if **o* was followed by a lost “glottal fricative” **h* (= **h₂*), while he rejects all other compensatory lengthenings caused by laryngeals. This leads to unnecessary postulation of a glottal fricative for all cases of Indo-Iranian *ā* = European **o*, and his reconstruction methodology is very problematic in general.

⁴ Chronology and details are disputed, see Kloekhorst 2008a; 2008b: 98–9; 2014: 250, 553–9; 583–4 vs. Melchert 1994: 105, 131, 243–4, 264; 2012b.

unclear if this was an early change and if it happened in all Anatolian languages (in Lycian **e* and **o* clearly merged into *e*, but the quantity distinction was lost there).

The mechanism of this sound change is not really clear: could it have been a lengthening of “tense” [o] vs. “lax” [ɛ] (Keydana 2012)? Or is it rather a kind of relic of an originally long vowel (Kümmel 2012: 308–20), similar to what Brugmann proposed (cf. also Viredaz 1983: 35–7; Woodhouse 2012: 2 n. 1; 2015: 6–9)? This last option would presuppose a common innovation of most other languages, i.e. shortening of **ō* in most environments (preceding **oH > *ō*); however, this is difficult to reconcile with preserved IE **ō* in at least forms with lengthened grade.

14.2.4 The Vowel Merger

The most striking feature of Indo-Iranian is the merger of all non-high vowels instead of partial mergers in the neighbouring languages; elsewhere this is only found in Luwian (at least in Luwian). It is probable that this merger happened in two stages: first a lowering with a merger of non-front **o = *a > (back) *a*, then a merger of front **æ = *a > (central) *a*. The intermediate stage with **æ : *a* might be reflected by some Uralic loanwords, but this is not certain.

The more restricted merger of **a* and **o* is much more widespread: it is attested both in Anatolian (except Lycian but cf. above) and a “north-eastern European” area from Albanian and Messapic to Balto-Slavic and Germanic. In fact, only Tocharian and the southern languages from Celtic to Armenian show a distinction of these vowels. Thus the first step of the Indo-Iranian merger might be part of a larger areal development. For long **ā* and **ō* the merger is restricted to Anatolian, Germanic and Slavic, and in non-final syllables it is also found in Celtic. Albanian merges **ē* and **ā*, probably together with Messapic and Phrygian.

14.2.5 The Liquid Merger

The apparently complete merger of PIE **l = *r > *r* is not found anywhere else in IE. Substrate influence is therefore quite probable, but no known language in the relevant regions shows this phenomenon. Note that the often assumed “retention” of *l* in some cases in Iln. languages is probably a mirage with no historical foundation (see Hock 1991: 138; Mayrhofer 2004); there is no attested variety in which *l* shows a statistically valid correlation with PIE **l*. Preservation is also contradicted by the fact that the liquid merger fed the *ruki* development, i.e., PIE **ls* turned into **rs > *rś* in all of Indo-Iranian, cf. the

root **k^wels-/k^wols-/k^wls-* (Gr. *τέλσον* ‘furrow’) > PIIrn. **kārš-/karš-/kṛš-* ‘to pull, draw, plough’ > Ved. *carṣ-/karṣ-/kṛṣ-* = Iranian **karš-/kərš-*.⁵

14.2.6 Weak Stem in Accusative Plural

While in Indo-Iranian the accusative plural belongs to the “weak” stem, elsewhere it normally belongs to the “strong” stem. The only exceptions are “proterokinetic” *i/u*-stems with **-ej-es* : **-i-ms*; **-ew-es* : **-u-ms*. The simplest explanation for this difference is an Indo-Iranian innovation, used to repair the homophony of accusative **-m̥s* > *-as* = nominative **-es* > *-as*, building on the existing model of the *i/u*-stems (see Hock 1974).

14.2.7 Laryngeal Aspiration

Indo-Iranian is the only branch with incontestable examples of aspiration caused by a following laryngeal. The most famous examples show **h₂* after stops:

- **meǵχ-* > **majh-* > **maj^hh-* ‘big’ > Ved. *máh-*, cf. Gr. *μέγα*, Hitt. *mēkk-* (together with Iranian **majh-* > **mach-* > **mac-* > *mas-*, *maθ-*, see Section 5.3)
- **sístχa-* > **sištha-* > Ved. *tiṣtha-* ‘to stand’, cf. Gr. *stā-*, *-sth-* (in cases like *Ὀρεσθ-*)
- **p_ltχú-* > **prthú-* ‘broad’ > Ved. *prthú-* = Av. *pər^θθu-*, cf. Gr. *Πλαταιαί*
- pf.2sg. **-tχa* > Ved. *-tha* = Av. *-θa*, cf. Gr. *-tha*, cf. also Vedic mid.2sg. *-thās*. For **h₁* this is controversial, but there are some potential examples:
- **pónt-e/oh-* ~ **p_{nt}-h-* > **pántā-* ~ **path-* > Av. *pañtā-* ~ *paθ-* (see de Decker 2012)
- 2pl. **-the* > Ved. *-tha* = Av. *-θa* (but cf. Sabellic **-tā* < **-tah₂* if not from the dual?).

Continuation of **h₁* as an aspirating sound would also be supported by **d^hed^hh-* > **dadh-* > **dath-* > Proto-Iranian **daθ-* (see above), but this example does not show secondary aspiration as such.

⁵ There have been attempts to include pre-PIIrn. **l* as [+high] (which would require a change to a palatal or at least retroflex) in the sounds that triggered *ruki* (see Lipp 2009, 1: 33 n. 72) but this is contradicted by its different behaviour in all other *ruki* languages. There is no other evidence that IE or pre-PIIrn. **l* had an articulation place different from **n*. Fortunatov 1881 claimed a special development of **lt* > Skt. *ṭ* (etc.), but this is generally rejected today; Pyysalo (2013: 227–43) has tried to modify Fortunatov’s Law by also including **r* but assuming an adjacent “diphonemic pair” **ah/ha* = laryngeal as additional conditioning. This cannot be accepted since it is phonetically unmotivated, and the general approach is based on a flawed reconstruction methodology and much dubious material.

In Greek only $*T\chi V > T^hV$ seems to be possible, but this is disputed (cf. Cowgill 1965 vs. Peters 1991), and other branches show no clear evidence. Armenian and Slavic seem to show $x < *k\chi$, cf. $*\hat{t}k\acute{\alpha}\chi k\chi\text{-}/(t)\hat{k}\chi k\acute{\alpha}\chi\text{-}$ ‘branch’ > Arm. $c'ax$, CSL. $*sox\grave{a}(c)$ = Ved. $\acute{s}\acute{a}kh\bar{a}$ -, Sogd. $\acute{s}\bar{a}x$ (beside MPers. $\acute{s}\bar{a}g$), but this does not necessarily presuppose an intermediate stage with aspiration. No other evidence is found in languages without phonological aspiration.

Notably, it is not altogether clear if Iranian participated in the development of aspiration, or if clusters of stops + $*h$ just underwent preconsonantal fricativization of stops followed by loss of $*h$ (see Kümmel 2018c: 162–4).

14.2.8 A Striking Difference

There is one striking difference between Ilrn. and the rest of Nuclear IE (= Indo-Tocharian, see Olander 2019):⁶ “vocalization” of laryngeals leads to low(er) vowels everywhere from Tocharian to Celtic, and from Greek to Germanic, but in Indo-Iranian, we only find the high vowel i , and Iranian and Indo-Aryan do not agree in the conditioning, with Iranian most often showing no vowel. The simplest explanation for this situation is that epenthesis was partly post-PIlrn. (see Kümmel 2016c; Aufderheide & Keydana 2016), and that i is not a direct reflex of the laryngeal. It can thus rather be compared to Greek cases of “schwa secundum” = i insertion (de Vaan 2009). This rather strong difference might be interpreted as an early divergence of Indo-Iranian vs. the rest. However, differences in details exist between all other branches, too, so it remains unclear how fundamental this is.

14.3 The Internal Structure of Indo-Iranian

In the oldest stage, there are no fundamental or significant grammatical differences between Iranian and Indo-Aryan. The morphology and syntax of the earliest Vedic and Old Avestan texts are very close, and the main differences are found in phonology and lexicon.

14.3.1 Phonological Features

An overview of the main phonological differences is shown in Table 14.1 (clear innovations are shaded).

⁶ There are no really good examples of “vocalization” in Anatolian: weak stems like $as\text{-}$, $ad\text{-}$ for $*h_1s\text{-}$, $*h_1d\text{-}$ are possibly analogical, and Luw. $tuwatr\text{-}$, Lyc. $kbatra$ ‘daughter’ is not clear enough.

Table 14.1 *Main phonological differences between Iranian and Indic*

Proto-Indo-Iranian	Iranian	Indic	Remarks
<i>*b, *d, *g : *b^h, *d^h, *g^h</i>	<i>b, d, g</i>	<i>b, d, g : b^h, d^h, g^h</i>	merger
<i>*p, *t, *k / _C</i>	<i>f, θ, x</i>	<i>p, t, k</i>	fricativization
<i>*ph, *th, *kh</i>	<i>f, θ, x</i>	<i>ph, th, kh</i>	(only a special case of the previous row)
<i>*ć, *j</i>	<i>*ts, *dz > s/θ, z/ð</i>	<i>ś, j</i>	depalatalization
<i>*j, *j^h : *ǵ, *ǵ^h</i>	<i>*dz : *j</i>	<i>j, h : j, h</i>	merger
<i>*s</i>	<i>h</i>	<i>s</i>	
<i>*š</i>	<i>š</i>	<i>ś</i>	only phonetic
<i>*zD, *zD</i>	<i>zd, žd</i>	<i>:d, :d</i>	not yet in WIA
<i>*tst, *dzd^h</i>	<i>st, zd</i>	<i>tt, dd^h</i>	different simplification
<i>*tš : *kš</i>	<i>*č > š : xš</i>	<i>kṣ : kṣ</i>	dissimilation, merger
<i>*r</i>	<i>*ər (?)</i>	<i>*ṛ</i>	only phonetic?
<i>*ər</i>	<i>ar (~ ər?)</i>	<i>ṛr/ūr</i>	see Cantera 2001
<i>*h-</i>	<i>h/x ~ Ø</i>	<i>Ø</i>	Kümmel 2016a : 83; 2018c : 166
<i>*Dh, *Bh, *Jh *Dahi/u</i>	<i>*θ, *f, *ts *θai/u</i>	<i>d^h, b^h, h *dai/u</i>	Kümmel 2016a : 82–3; 2018c : 165–6
<i>*-CHC-</i>	<i>CC</i>	<i>CiC</i>	see Werba 2005 ; Kümmel 2016c : 219–22
<i>*pt-</i>	<i>ft</i>	<i>pit</i>	epenthesis (Kümmel 2016c : 222–3)
<i>*pst-, *db-, *dm-</i>	<i>fšt, db, dm</i>	<i>st, b, m</i>	see Kümmel 2014 : 211–12
<i>*-kš(t), -kšt-</i>	<i>xš(t), xšt</i>	<i>k, kt</i>	simplification, see Kümmel 2014 : 212–14

For Proto- or Common Iranian affricates see [Lipp 2009](#), 1; 183–91; [Peyrot 2018](#); for the development of “thorn” clusters (**tḱ > *tć > *tš* etc.) see [Lipp 2009](#), 2: 1–313 with refs.

14.3.2 *Morphosyntactic Features: Iranian vs. Vedic*

There are a number of mostly minor differences in morphological detail between Old Iranian and Vedic Sanskrit. Most often, Indo-Aryan has innovated while the older stage is better preserved in Iranian ([Table 14.2](#)).

However, there are also some cases where Old Avestan stands against an innovation in Younger Avestan, Old Persian and Vedic, so it seems that there was a parallel development in Indo-Aryan and younger Iranian ([Table 14.3](#)).

Only rarely is it Old Avestan that innovates vs. archaisms in Younger Avestan and (if applicable) Vedic ([Table 14.4](#)).

Table 14.2 *Morphological differences between Iranian and Indic*

	Iranian	Indic	Remarks
gen. : loc. dual	*-ās : *-aw	*-awš	cf. Slavic *-u < *-au(š)
instr.-dat.-abl. dual	*-ayb ⁶ yā > OAv. -ōibiiā, YAv. -aēbiia, OPers. -aibiyā	*-āb ⁶ yā(m) > -ābhyām	
u-stem type -āw- (nom. sg., acc.sg.)	-āuš, -ām	-úš, -úm	
n. n-stem gen.sg.	*-ans > OAv. -āng	-nas	
a-stem instr.sg.	*-ā	-éna (-ā)	
comparative *-yās-, perf.ptc. *-wās-	*-yāh, *-wāh-	-yāms-, -vāms-	
nt-ptc. to thematic stems	*-ant-	-at-	ablaut taken over from athematic bases
1sg. pronoun gen.	*mana	māma	but cf. Khot. <i>mamā</i>
2pl. pronoun nom.	*yūž-am	yūyām	contamination with 1pl. <i>vayām</i>
3ps. encl. dative possessives	*hai ~ *šai av. <i>ma-</i> , <i>θβa-</i>	— —	loss in Indic loss in Indic (but also in later Iranian)
distal demonstrative	*awá- acc.sg. *aw-ām	amú- acc. sg. *am-ú	see Klein 1977
interrogative numeral 'one'	ci-, ca- : ka-	ká-(kím)	generalization of k-
middle thematic ptc.	*aywá-	*áyka-	
active optative	-ī- : -yā-	-māna-	but cf. MIA -mīna-
3pl. SE	-at (: -rš)	only -yā- only -ur	(few relics of * ^o aH-ī-)
subj.mid.1sg.	*-ānai (~ *-āi)	only *-āi	
mid.3pl.	*-ārai, *-āra(m) ~ *-rai, *-ra(m)	only *-rai	

14.3.3 The Special Case of Nuristanic

The so-called Nuristani languages are spoken just between Eastern Iranian and NW Indo-Aryan in the Hindukush region. They are only attested in modern times and represent a group of transitional languages between Indo-Aryan and Iranian, rather difficult to classify due to the lack of ancient data. In some features, they agree with Iranian, in others with Indo-Aryan, but they clearly differ from both since early times:

Table 14.3 *Morphological archaisms in Old Avestan*

	Old Avestan	Elsewhere	Remarks
accusative 1/2pl.	$nā < *nās$ $vā < *wās$	<i>*nas</i> , <i>*was</i> (= dative-genitive)	cf. Lat. <i>nōs</i> , <i>uōs</i> , OCS <i>ny</i> , <i>vy</i>
nom.acc.pl.n. <i>r/</i> <i>n</i> -stems	<i>-ār</i> ³	YAv. <i>-qn</i> = Ved. <i>-ān-i</i>	cf. Hitt. <i>-ār</i>
1sg. present	<i>-ā ~ -āmi</i>	only <i>-āmi</i>	cf. general European <i>*-ō</i>
velar ~ palatal alternation	<i>aōgō</i>	YAv. <i>aōjō</i> = Ved. <i>ójas</i>	generalized velar in Ved. <i>ágas-</i> , <i>ókas-</i> , elsewhere palatal
inflection of <i>*wicwa-</i> 'every, all' <i>*anya-</i> 'other'	OAv. <i>vīspāñhō</i> "Median" <i>aniyāha</i>	Ved. <i>vísve</i> , anyé YAv. <i>vīspe</i> , <i>ahiie</i> OPers. <i>aniyai</i>	pronominal desinences of adjectives (archaism in OAv. not sure)

Table 14.4 *Morphological innovations in Old Avestan*

	YAv. (= Ved.)	Old Avestan	Remarks
gen.sg. <i>*krátwas</i> , <i>*pačwás</i> , <i>*pitvás</i>	<i>xraθβō = krátvas</i> <i>pasuuō = pašvás</i> <i>*piθβō = pitvás</i>	<i>xratēuš</i> <i>pasēuš</i> <i>pitēuš</i>	most productive inflectional type
acc.pl. <i>*pṛtwás</i> weak stem <i>*majh-</i> , <i>*dadh-</i> > <i>*mac-</i> , <i>*daθ-</i>	<i>pərəθβō</i> <i>mas-</i> , <i>daθ-</i> = <i>mah-</i> , <i>dadh-</i>	<i>pərətūtš</i> <i>maz-</i> , <i>dad-</i>	analogy after strong stem <i>mazā-</i> , <i>dadā-</i>

- "Iranian" features: depalatalized **ts*, *dz* distinct from **č*, *ǰ*; no aspirates (= deaspiration)⁷ – rather trivial developments (also attested in neighbouring Indo-Aryan but much later)
- "Indic" features: **tst*, *dzdʰ* > *tt*, *dd*; **ər* > **i/ur*;⁸ preserved *s*, no fricativization
- special features:
 - **čš/tč* > **tš* > **ts* vs. Iranian *š*, Indic *kš*, cf. Kati *ič* 'bear'

⁷ However, since Dameli (in spite of some doubts) probably belongs to Nuristanic and appears to show voiceless aspirates in line with Indo-Aryan, the loss of voiceless aspirates in the rest of Nuristanic may be a late innovation. For voiced aspirates, the merger of the palatal aspirates with the simple voiced palatals presupposes a chronology different from Indo-Aryan, but this only requires that aspiration was lost before the debuccalization of palatal aspirates.

⁸ With one probable exception: **wərnā-* 'wool' did not become **wurnā-* (> Ved. *úrñā-*) but **warnā-* > **wārā-*; cf. Av. *varənā-*.

Table 14.5 *Phonological changes in Iranian, Nuristanic and Indic*

	*ć, *j, *j ^h	*d ^h	*d	*ǎ	*tst	*s	*tr	*th	*tš	*st	*rn	*nt
Irn.	ts > s/θ, dz > z/d	d	d	a	st	h	θr	θ	š	st	rn (> rr/ṛṇ)	nt > nd
Nur.	ts, dz	d	d	i/u	tt	s	tr	t	ts	št	r	nt (> t)
Ind.	ś, j, h	d ^h	d	i/u	tt	s	tr	t ^h	kṣ	st	rṇ (> ṇṇ)	nt > NW nd



Figure 14.1 The Indo-Iranian languages

- *st > št (Kati *dušt* ‘hand’); *š/s > s (secondary, see Cathcart 2011); *Vrn* (> *rr?) > *Ṽr*
- no voicing in *nt*, *nk*, *nč* (vs. most neighbours).

See Table 14.5 (innovations shaded).

The most recent discussion is by Werba 2016, who argued that Nuristanic forms a subgroup with Indo-Aryan; but even if he was right to stress that similarities to Iranian do not require a common stage, the differences from Indo-Aryan are strong enough that for all practical purposes, Nuristanic has to be treated as an independent third branch (see Figure 14.1). It did not participate in most early innovations of either Iranian or Indo-Aryan.

In the lexicon, Nuristanic shows some possibly ancient similarities to Iranian (e.g., **khandā*- ‘to laugh’, **wainā*- ‘to see’, **arjana*- ‘millet’, **pragāma(ka)*- ‘young animal’, **jāyan*- ‘winter’, **tridaća* ‘13’, **katrudaća* ‘14’), but much more often it agrees with Indo-Aryan, which, however, could be due to secondary influence in most cases. It does not share most typical early Iranian (potential) innovations like **g^hauša*- ‘ear’, **kaśman*- ‘eye’, **wasunī*- ‘blood’, **ātr*- ‘fire’, **swar*- ‘to eat’.

14.3.4 Lexical Differences

Some examples of lexical differences between the main branches are shown in Table 14.6 (dating of innovations is of course uncertain in Nuristanic due to the lack of ancient data).

Table 14.6 *Examples of lexical differences between Iranian, Nuristanic and Indic*

	Iranian (Avestan)	Nuristanic	Indic	Remarks
‘fire’	<i>ātar-</i> (-ayni-)	<i>*angāra-</i>	<i>agnī-</i>	choice of inherited terms; replaced by <i>angāra-</i> ‘glowing coal’ in Nuristanic and “Dardic” IA
‘water’	— <i>āp-</i>	— <i>*āp-</i>	<i>vār/udán-</i> <i>āp-</i>	derivatives in Irn. Nur.
‘rain’	<i>vāra-</i>	<i>*warša-</i>	<i>varṣá-</i>	derivative of <i>*waHr</i> ‘water’
‘eye’	(<i>aši</i>) <i>cašman-</i>	<i>*akši</i>	<i>ákṣi</i> <i>cákṣuṣ-</i>	parallel innovations
‘ear’	<i>*uši</i> <i>*gauša-</i>	<i>*karna- ></i> <i>*kāra-</i>	<i>kárṇa-</i>	cf. Av. <i>kar²na-</i> = Ved. <i>karṇá-</i> ‘deaf’ Ved. <i>ghóṣa-</i> ‘sound’
‘to eat’	<i>xʷar-</i>	<i>*yaw-</i>	<i>ad-</i>	<i>*yaw-</i> also in Waxi and Chitral IA
‘to drink’	<i>xʷar-</i>	<i>*pā-</i>	<i>pā-</i>	relics in easternmost Iranian: Waxi <i>pəv-</i> < <i>*piba-</i>
‘to see’	<i>vāena-</i>	<i>*waina-</i>	<i>pásya-</i>	Ved. <i>véna-</i> ‘to look after’ ⁹ Av. <i>spasiti-</i> ‘to watch’
‘blood’	<i>vohunī-</i>	<i>*asan-</i>	<i>ásṛk, asan-</i>	
‘bird’	<i>vi-</i> <i>mərəṣya-</i>	? <i>*mṛga-</i>	<i>ví-</i> <i>pakṣin-</i>	Ved. <i>mṛgá-</i> ‘animal, game, deer’
‘spring’	<i>*wasar-</i>	<i>*wasanta</i>	<i>vasantá-</i>	
‘winter’	<i>zim-</i> <i>zaiian-</i>	<i>*jayan-</i>		
‘ice’	<i>aēxa-</i>	<i>*yaja-</i>	<i>hemantá-</i> ?	
‘snow’	<i>snaiga-</i> <i>*jim-</i> <i>vafra-</i>	<i>jima-</i>	<i>*sniḥ-, *sneha-</i> <i>himá-</i>	
‘moon’	<i>māh-</i>	<i>mās-</i>	<i>más-</i>	
‘sky’	(<i>diiu-</i>) <i>asmān-</i> <i>abra-</i>		<i>candrámās-</i> <i>dyāv-</i>	
‘stone’	<i>asan-,</i> <i>asəṅga-</i> <i>*warta-</i> <i>*gari-</i>	<i>*warta-</i> <i>*giri-</i>	<i>ásman-, áśn-</i> <i>*warta-</i>	
‘mountain’	<i>gaʹri-</i> <i>paʹruuata-</i> <i>*kaufah-</i>	<i>*dārā-</i>	<i>giri-</i> <i>pārvata-</i> <i>*dhārā-?</i>	

⁹ Derived from the noun **wainá-* > Ved. *vená-* ‘watcher’, MPers. *wēnag* ‘guard, watchman’; Indic preserves the narrower meaning; the broadened meaning may also be reflected by an apparently old loanword into Western Uralic, cf. **wajna-* > Southern Saamic **wuojnē-* ‘to see’, Mordvin *vano-/vanō-* ‘to watch’ (see Holopainen 2019: 312–13).

For differences in most of the agricultural terminology (as opposed to animal husbandry), see Kümmel 2017.

14.4 The Relationship of Indo-Iranian to the Other Branches

14.4.1 The Central IE Sound Shift

Indo-Iranian seems to belong to the group of IE languages that reflect voiced aspirates and thus presuppose the “central IE sound shift” (Kümmel 2012: 304–6; 2016c: 130–2), i.e. a chain shift from PIE (PIA) $*d : d^h > \text{Central IE } *d^h : d$. This is clear for Indo-Aryan, which has had breathy voiced stops ever since Sanskrit. However, it has been proposed that this change did not happen in Iranian (and Nuristanic) where aspiration of media aspirata (MA) is not directly preserved (Lubotsky 2018), so the sound shift would only be an Indo-Aryan innovation, parallel to Greek etc. This is not very easy to determine. One possible argument for pre-Iranian aspiration might be Bartholomae’s Law, the outcome of which is still faithfully observed in Old Avestan. However, this law is possibly older still, since it works even better with pre-shift phonology (cf. progressive voicing as in Turkish) if implosives did not participate in the voicing distinction (cf. above). Thus, its reflection in Old Avestan does not necessarily presuppose aspiration but only some distinction between “media” and “media aspirata”. At first sight, Iranian $*dugdar-$ ‘daughter’ $< *dugd^h ar-$ appears to presuppose a post-PIIrn. application of BL, since $*dugd^h tar-$ can only have arisen secondarily by loss of the laryngeal in $*dughtár- < *d^h ugxtér-$. However, such an allomorph might already have been present in PIIrn. and simply been ousted in Indic (see Lipp 2009, 2: 370–84; Kümmel 2018c: 169). Within a “glottalic” reconstruction of PIIrn., one could also assume $*dugHtar- [ʔgʔ] > *dug(H)tar- [gʔ] > *dugdar-$ so that we would not strictly need aspiration to be present. However, there is at least one change in Iranian that seems to presuppose aspiration of the MA, namely the transfer of postnasal aspiration to the preceding onset seen in $*teng^h- > *tang^h- > \text{Iranian } *thang- > *θang-$ ‘to pull’ and maybe also in $*kumb^h a- > *khumba- > \text{Iranian } *khumba- > *xumba-$ ‘pot’. This might be supported by a systemic argument: Indo-Iranian does not show any bias against “mediae” after nasals, as one might expect for implosives, so it seems more probable that the “mediae” had already become voiced explosives.

14.4.2 The Satem Phenomenon

The so-called *satem* languages show palatal or depalatalized coronal affricates or fricatives corresponding to *centum* velar stops, and simple velars corresponding to *centum* labialized velars (labiovelars). In a third type of

correspondence, all languages have simple velars. The usual PIE reconstruction is so-called “palatals” in the first case, “labiovelars” in the second and “pure velars” in the third. However, the existence of real “pure velars” in PIE has been questioned, and this type of correspondence could also be explained by neutralization of an original twofold contrast between “palatovelars” and “labiovelars”.

The *satem* languages comprise all Eastern languages except Tocharian, while the areal distribution of *centum* languages looks much less compact, including the outliers Anatolian and Tocharian, and the European West and South. Therefore, the *centum* situation is most probably original, and the *satem* group underwent a chain shift $*k^w : *k > *k : *c$. This is a rather trivial phonetic change, but details of phonologization and distribution are far from trivial, cf. forms like $*(H)o\acute{k}t\acute{o}H$ ‘eight’, synchronically isolated. This requires the assumption of one areal change, possibly cutting across other isoglosses.

The satemization is apparently connected to another areal feature, that of the *ruki* rule, i.e. a retraction of $*s$ after non-anterior sounds, which is found in more or less the same branches, though to different degrees (with some restrictions in Slavic and Baltic, and only to a very limited extent in Armenian and Albanian, see Martirosyan 2010: 709–10 with refs.). This allophony may have been more widespread in IE but was only phonologized in *satem* languages since only these developed additional sibilants from other sources (see Andersen 1968).

Similar developments of “palatals” are found in Luwic Anatolian, but then combined with preserved labiovelars. According to the most recent investigation (Melchert 2012a), there was a conditioned palatalization of old “palatals” only; but the claim that original “pure velars” contrastively remained unpalatalized is unsubstantiated: the only example of a preserved velar before a front vowel is Luwian $k\acute{s}\check{a}(i)$ - ‘to comb’, and this may have analogical k - or even continue $*ks$ - (there was a regular change of $*ks > kis$ in Hittite, no counterexamples in Luwian). So Luwian might in fact reflect the usual “*centum*” merger of “palatals” and “velars”, followed by a conditioned palatalization of the resulting velars. However, some words appear to show Luwic “palatals” in environments where secondary palatalization would be improbable: cases like Luw. *zanta* ‘down’ (Goedegebuure 2010) $< *k\acute{a}nt$ - (cf. Hitt. *katta*, Gr. *κατά*) and also HLuw. *azu(wa)*- ‘horse’, *zuwan*- ‘dog’ $< IE *ekw(o)$ -, $*kwon$ -, if the latter are not to be read as *asu(wa)*-, *suwan*-, borrowed from WIA (as argued by Szemerényi 1976; Lipp 2009, 1: 273–302). If these words show a genuine Luwic development, this looks much more like preserved IE “palatals” than anything secondary.¹⁰ Interestingly,

¹⁰ One might consider an intermediate stage with a secondary front vowel in the first case, so something like pre-Luwic $*k\acute{m}t^{\circ} > *kent^{\circ} > *k\acute{a}nt^{\circ} > zant^{\circ}$, but this does not work for the two words with $*k^w$.

recent research has also found some *ruki*-like developments in Luwian (Rieken 2010), which would support the idea that the Luwic developments are *satem*-like. Currently, it is still unclear how exactly this might be explained.

14.4.3 Middle Primary Endings

The “primary” endings of the middle are marked by **-y*, identical to **-i* used in the corresponding endings of the active. Here IIrn. agrees with Armenian, Albanian, Greek and Germanic, while the more “peripheral” branches Anatolian, Tocharian, Italic and Celtic show **-r*. The latter has been interpreted as an archaism and marking by **-i/y* as analogical (see Dunkel 2014: 669–70). However, much is still unclear here. In Phrygian, we find *-toy* earlier than *-tor* (but never *-to*). In Tocharian, the preterit middle 1sg. **-ai*, 2sg. **-tai* could be explained as relics of older *-i*-endings (see Malzahn 2010: 44–6 with refs.). In Celtic and Italic, *-r* is not used in all cases, which might point to an incomplete spread.

In Greek, the 1pl. and 2pl. endings are not marked by *-i* (mirroring the situation in the active), but in Indo-Iranian, they also have a final diphthong **-ay*, resulting from a further spread, viz. 1pl. **-mad^hay* < **-med^hoj* for **-mesd^hχ*. The same probably happened in Armenian, Albanian and Germanic (see Kümmel 2018b: 194).

14.4.4 Verbal Dual Endings

The non-present endings Ved. 2du. *-tam*, 3du. *-tām* seem to agree perfectly with Gr. 2du. *-ton*, 3du. *-tān* < **-tom*, **-tām*. However, the corresponding Avestan endings *-təm* and *-tqm* are both used for the 3du., and Toch.B 3du. *-te-m* (with a secondary nasal) might support the use of **-tom* for the 3sg. Similarly, Avestan does not reflect the distinction of Ved. 2du. *-thas* : 3du. *-tas* but used *-θō* = *-tō* indiscriminately. Gothic 2du. *-ts* seems to agree, but Greek uses a different ending with no distinction 2=3du. *-ton*. The Baltic 2du. **-tās* and Slavic 2du. *-ta*, *-te*, 3du. *-te* do not agree completely, so a precise reconstruction remains difficult (Pooth 2011 has argued for a secondary differentiation and a connection to the middle).

14.4.5 Formation of Accented Personal Pronouns

The PIIrn. stems of the accented non-singular personal pronouns are 1pl. **as-má-*, 2pl. **uš-má-* < **ṇs-mé-*, **us-mé-* vs. 1du. **āwá-* < **aH-wá-* < **ṇH-wé-* (2du. **yuwá-* < **uH-wá-*). This agrees most closely with Greek 1pl. **ahme*, 2pl. **uhme* > Aeol. *ἄμμε*, *ὕμμε*; Dor. *ἄμε-*, *ὕμε-*; Ion.-Att. *ἡμε-*, *ὕμε-* and 1du. **nō-we* < **noH-we* (but 2du. *σφω*). Elsewhere we either find only **nōs*, **wōs*

(Italic, Balto-Slavic, Albanian) or 2pl. **uswe*: Celtic 2pl. **swīs*; Germanic **izwiz* or even 1pl. **ŋswe* > Hitt. *anze-*, *sume-*, Luw. *anzu-*, *unzu-*. The PIE situation is not very clear: apparently extension of the base by both **-me* and **-we* was possible, and various scenarios have been proposed:

- a. pl. **-me* vs. du. **-we* (Cowgill 1965 = Ilrn. + Gr. Archaism)
- b. 1st **-me* vs. 2nd/3rd **-we* (Katz 1998: 279)
- c. “inclusive” **-me* vs. “exclusive” **-we* (Dunkel 2014: 494, 499, 569–74).¹¹

An original inclusive/exclusive distinction appears most promising, but typologically, an inclusive first person (in the usual definition ‘me and you’) often shows a marker of the second person, and this might favour a distribution of first person exclusive **-me* (cf. 1sg. **me-*) ‘me and someone else but not you’ vs. first person inclusive ‘me and you’ + second person **-we* (cf. second person **wo-*). In this case, Greek and Ilrn. would show a common innovation, i.e. generalization of the exclusive marker **-me* in the first person plural followed by its spread to the second person plural, and generalization of the inclusive marker in the first dual. However, this innovation need not be exclusively Greek and Ilrn., since corresponding forms might have been lost in all branches that lost these extended forms, i.e. Italic, Albanian, Balto-Slavic and Tocharian.

14.4.6 *Augment*

The so-called augment, i.e. a verbal prefix marking the past vs. the injunctive is only found in Indo-Iranian, Greek, Armenian, Phrygian and Albanian and might be either an archaism lost elsewhere or a common innovation. However, it seems clear that much of the development was parallel rather than shared, since in the earliest records, the prefix had not yet become an obligatory marker. Therefore, the original situation must have been a much less grammaticalized item, in which case it is much easier to assume its loss in other branches.

14.4.7 *Primary Superlatives*

The primary superlative is derived from the primary comparative by the suffix **-t(H)o-* in Indo-Iranian, Greek and Germanic, while Italic and Celtic show **-is-m(H)o-*. Since both suffixes correspond to some original numerals (see Luján 2019), a parallel development is not unlikely.

¹¹ Dunkel’s reconstruction is based on the particles **me* ‘within, together with’ and **we* ‘or’, and he uses an unusual definition of inclusive = ‘me and a third party’ vs. exclusive = ‘me without a third party’.

14.4.8 *Secondary Comparatives*

The suffix **-tero-* serves as a productive secondary comparative only in IIrn. and Greek, while elsewhere it can only be derived from pronouns and adverbs. However, the corresponding superlative formation is different: Greek *-tato-* vs. PIIrn. **-tama-*. Therefore, the development was not identical, so the probability of a parallel extension of the existing departicular system is quite high.

14.4.9 *Formation of Decades*

The PIIrn. cardinal numerals ‘thirty’, ‘forty’ and ‘fifty’ are formed by a suffixoid **(d)ća(n)t-*, based on compounds with **-d̥kōmt-/d̥k̑mt-*. This seems to agree only with Celtic, where all decades from thirty to ninety are formed with **-d̥kōmt-*. By contrast, Armenian, Greek, Italic and Tocharian show a slightly different formation with cardinal + collective **d̥kōmtχ/d̥k̑mtχ*, and Germanic and Balto-Slavic only use a syntagma with the free word **d̥k̑mt-* (cf. Rau 2009 for an overview and discussion). Since the most original situation remains unclear, the significance of the Celtic–IIrn. agreement is unclear.

14.4.10 *Instrumental, Dative and Ablative Dual and Plural*

In endings of the instrumental, dative and ablative dual and plural, the PIIrn. set **-b^hyā*, **-b^hiš*, **-b^hyas* corresponds more closely to the “southern” set **-b^hoH*, **-b^hiś*, **-b^hos* attested from Armenian to Celtic, in contrast to “northern” endings with **-m^o* in Germanic and Balto-Slavic. Both sets are probably innovations, but the precise development still needs to be clarified (see Melchert & Oettinger 2009; Kim 2013); in any case, the agreement with the southern group indicates closer contact, but differences in details favour an areal development rather than an inherited innovation from a common pre-stage.

14.5 **The Position of Indo-Iranian**

There can be no question that all Indo-Iranian languages are related to one another much more closely than to any other IE language, so Indo-Iranian is clearly defined as a primary branch of IE. The relationship of Indo-Iranian to other branches, however, is much less easy to describe. It has variously been grouped together with quite distinct branches in the history of IE linguistics.

14.5.1 *Different Trees*

Nearly all cladistic models assume Anatolian to have split off first (“Indo-Hittite” model) from PIE with the remaining branches becoming NIE, and most

also assume a second split-off of Tocharian vs. Inner IE (= Indo-Celtic, see [Olander 2019](#)) from NIE. Otherwise, they differ in many ways, as in the following overview, with the branches grouped according to how close they are to Indo-Iranian:

- Schleicher's first trees (1860; 1861; 1862): 1. Graeco-Italo-Celtic, 2. Germanic-Baltic-Slavic
- [Gamkrelidze & Ivanov 1995](#): 1. Armenian, 2. Greek, 3. Germanic-Baltic-Slavic, 4. Italic-Celtic-Tocharian
- [Hamp 1990](#): 302: 1. Indo-Iranian = "Asiatic IE" vs. 2. "Residual IE" (all the rest including Tocharian)
- Starostin 2004 (core lexicon only, glottochronology):¹² 1. Balto-Slavic, 2. Germanic-Italic, 3. Armenian, Greek, Albanian

Trees based on computational phylogenetic methods:

- [Ringe, Warnow & Taylor 2002](#) (mixed features; Germanic not classified): 1. Baltic-Slavic, 2. Greek, Armenian, 3. Italo-Celtic, 4. Albanian
- [Gray & Atkinson 2003](#); [Bouckaert et al. 2012](#) (core lexicon only, problematic database, Bayesian): 1. Albanian, 2. Baltic-Slavic-Germanic-Italic-Celtic, 3. Greek-Armenian
- [Chang et al. 2015](#) (same database and method, different calibrations): 1. Baltic-Slavic-Germanic-Italic-Celtic, 2. Greek, Armenian, Albanian.¹³

Thus all neighbouring sub-branches except Tocharian have been assumed to be nearest to Ilrn. In what follows, some important isoglosses are briefly discussed.

14.5.2 *Irrelevant Features: Shared Archaisms*

Many common features of Greek and Indo-Iranian are archaisms due to earlier attestation of these branches already in the second millennium vs. all other NIE branches. For example, preservation of:

- perfect as a distinct category
- original simple imperfect (vs. renewed marked formations in Tocharian, Armenian, Italic, Slavic)
- subjunctive and optative (vs. loss of optative in Celtic, Armenian, of subjunctive in Germanic, Baltic-Slavic)
- vocabulary and poetic language.

It is clear that such evidence is not relevant for subgrouping.

¹² Sergej Starostin, 2004, Handout, Workshop on the Chronology in Linguistics, Santa Fe.

¹³ This is also the result of the most recent application of Bayesian methodology based on a strongly improved new database in Jena (IE-CoR, with my own participation). The best-supported tree configuration still shows Indo-Iranian nearer to a group comprising Balto-Slavic and Italic-Celtic-Germanic than to Greek, Armenian and Albanian, but all this with very low certainty.

14.5.3 *Archaisms Shared with Anatolian (but not Greek)*

Some other archaisms are shared with Anatolian but not Greek. The clusters **tst* etc. were preserved in PIIrn. (> IA. + Nur. (?) **tt*, Irn. **st*, as elsewhere in Eastern IE). Morphological archaisms are the middle 3sg. ending **-á(y) < *-ó(-)* etc. and the active 3sg. ending *-s* (see Melchert 2015: 129–31; Kümmel 2018a: 245–52; 2018b: 1912–14); maybe also the numeral **syá-* ‘one’ (Kümmel 2016b) = Hittite *sia-/sie-* (but possibly also in Toch.B *še*, see Pinault 2006).

Notably, the preservation of consonantal laryngeals seems to be better than anywhere else in NIE:

- hiatus in Old Avestan and (less reliably) Vedic: e.g., subjunctive *dāt̪* {daat} = *dhāt̪* {d^haat} < **d^há(h)at*
- some laryngeals survived as some kind of **-h-* internally after stops into Iranian, causing devoicing of preceding obstruents (Kümmel 2016a: 82–3; 2018c: 164–5):
 - **majh-* > **mach-* > **mac-* > *mas-/maθ-* ‘great’ (vs. **majah-* > *mazā-*)
 - **dadh-* > **dath-* *daθ-* ‘put’ (vs. **dadah-* > *daḍā-*); **nābh-* > **nāph-* > *nāf-* ‘navel’ (vs. **nabah-* > *nabā-*)
 - **wabh-* > **waph-* > **waf-* ‘to weave’; **dahiwar-* > *dhaiwar-* > **thaiwar-* > **ṭaiwar-* ‘brother-in-law’
- *h-/x-* appears to be sporadically preserved in marginal Western Iranian (Kümmel 2016a: 83; 2018c: 166): e.g., MPers. *xīrs* ‘bear’, *xāyag* ‘egg’, *xāk* ‘dust’; *hēš* ‘ploughshare’, *hēsm/hēmāg* ‘firewood’, *hanzūg* ‘narrow’; Parthian *hand* ‘blind’. Especially the cases with *x-* can hardly be assumed to show a “prothetic” consonant. A similar case can be made for the eastern margin (Khotanese *h-*, see Kümmel 2020: 246)
- loss after *i/u* was probably only post-Proto-Iranian, cf. the contrast between lengthening and non-lengthening in cases like **wihrá-* > MPers. *wīr* vs. Sogd. *wīr-* ‘man’; **ǵīyvá-* > MPers. *zīw* vs. **žiwa-* > Sogd. *žaw-* ‘alive’; **duhrá-* > MPers. *dūr* vs. Khot. *dura-* ‘far’ (see Kümmel 2018c: 166–9).

14.5.4 *Unique Archaisms = Shared or Parallel Innovation Elsewhere*

Indo-Iranian exhibits a few unique archaisms that contrast with innovations elsewhere. For example, the middle 3pl. ending **-rá(y) < *-ró(-)* etc. which is not found anywhere else in the middle: all other branches including Anatolian generalized an ending containing **-nt-*. However, since the other branches do not agree in detail, this cannot be used as an argument for an early separation of Indo-Iranian vs. the rest. Two other morphological archaisms are the perfect 2pl. ending **-a < *-(H)e* and the preservation of a distinct genitive vs. locative dual only in Iranian, while all other branches either lack one or both of these

categories or show syncretism.¹⁴ In addition, there are numerous archaisms in the inflection of individual words and stems.

Recapitulating the phylogenetic relations of the Indo-Iranian branch, we may conclude the following:

- Indo-Iranian does not have a clear next relative.
- It is rather distinct in some respects, so an early split seems quite possible (Hamp's scenario), but only under the assumption of continued areal contact.
- There is good evidence for early proximity to Eastern Europe – with different developments shared with either the south (Greek, Albanian, Armenian) or the north (Baltic-Slavic, Germanic), or with the east (*satem* languages).
- An original position at the eastern fringe of Europe is corroborated by contacts with both Western and Eastern Uralic.

References

- Andersen, Henning. 1968. IE *s after i, u, r, k in Baltic and Slavic. *Acta Linguistica Hafniensia* 11, 171–90.
- Aufderheide, Tim Felix & Götz Keydana. 2016. Die i-Epenthese im Altindoeuropäischen: Laryngalkonsonantisierung oder "Verbindungsglied"? *Historische Sprachforschung* 129, 125–53.
- Bouckaert, Remco, et al. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337, 957–60.
- Brugmann, Karl. 1876. Zur Geschichte der stammbastufenden Declination. *Curtius' Studien* 9, 363–406.
- Cantera, Alberto. 2001. Die Behandlung der idg. Lautfolge (C)RHC- im Iranischen. *Münchener Studien zur Sprachwissenschaft* 61, 7–27.
- Cathcart, Chundra. 2011. RUKI in the Nuristani languages: An assessment. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 22nd annual UCLA Indo-European conference*, 1–11. Bremen: Hempen.
- Chang, Will, Chundra Cathcart, David Hall & Andrew Garrett. 2015. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91, 194–244.
- Cowgill, Warren. 1965. Evidence in Greek. In Werner Winter (ed.), *Evidence for laryngeals*, 142–80. London: Mouton.
- De Decker, Filip. 2012. Are Latin *pons*, *pontifex* and the Indo-European cognates evidence for an i-stem? *Journal of Indo-European Studies* 40, 11–45.
- Dunkel, George E. 2014. *Lexikon der indogermanischen Partikeln und Pronominalstämme*. Vol. 2. *Lexikon*. Heidelberg: Winter.
- Fortunatov, F. F. 1881. L + dental im Altindischen. *Bezenbergers Beiträge* 6, 215–20.
- Gamkrelidze, T. V. & V. V. Ivanov. 1995. *Indo-European and the Indo-Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture*.

¹⁴ There is no other certain trace of the perfect 2pl. ending *-a: Tocharian *-sa is not easy to derive from *-s-e, and the interpretation of Paelignian *lexe* as a 2pl. *leg-s-e is far from clear.

- With a preface by Roman Jakobson. English version by Johanna Nichols. Berlin: Mouton de Gruyter.
- Goedegebuure, Petra. 2010. The Luwian adverbs *zanta* ‘down’ and **ānni* ‘with, for, against’. In Aygül Süel (ed.), *Acts of the VIIIth international congress of Hittitology*. Vol. 1, 299–318. Ankara: Aralık.
- Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426. 435–9.
- Hamp, Eric P. 1990. The pre-Indo-European language of Northern (Central) Europe. In T. L. Markey & J. A. C. Greppin (eds.), *When worlds collide: The Indo-Europeans and the pre-Indo-Europeans*, 291–309. Ann Arbor, MI: Karoma.
- Hock, Hans Henrich. 1974. On the Indo-Iranian accusative plural of consonant stems. *Journal of the American Oriental Society* 94. 73–95.
- Hock, Hans Henrich. 1991. Dialects, diglossia, and diachronic phonology in Early Indo-Aryan. In William G. Boultz & Michael C. Shapiro (eds.), *Studies in the historical phonology of Asian languages*, 119–59. Amsterdam: Benjamins.
- Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic. Critical overview of sound substitutions and distribution criterion*. PhD thesis, University of Helsinki. <https://helda.helsinki.fi/handle/10138/307582>.
- Katz, Joshua T. 1998. *Topics in Indo-European personal pronouns*. PhD thesis, Harvard University.
- Keydana, Götz. 2012. Brugmann’s Law and the role of perception in sound change. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European: Papers in Indo-European phonetics, phonemics and morphophonemics*, 133–43. Munich: LINCOM.
- Kim, Ronald I. 2013. The Indo-European, Anatolian, and Tocharian “secondary cases” in typological perspective. In Adam I. Cooper, Jeremy Rau & Michael Weiss (eds.), *Multi nominis grammaticus: Studies in classical and Indo-European linguistics in honor of Alan J. Nussbaum on the occasion of his sixty-fifth birthday*, 121–42. Ann Arbor, MI: Beech Stave.
- Kiparsky, Paul. 2010. Compositional vs. paradigmatic approaches to accent and ablaut. In Stephanie W. Jamison, H. Craig Melchert & Brent Vine (eds.), *Proceedings of the 21st annual UCLA Indo-European conference*, 137–81. Bremen: Hempen.
- Klein, Jared S. 1977. The Indo-Iranian prehistory of the Sanskrit *asáu/amūm* pronoun. *Journal of Indo-European Studies* 5. 161–76.
- Kloekhorst, Alwin. 2008a. Čop’s Law in Luwian revisited. *Die Sprache* 46. 131–6.
- Kloekhorst, Alwin. 2008b. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- Kloekhorst, Alwin. 2014. *Accent in Hittite: A study in plene spelling, consonant gradation, clitics, and metrics*. Wiesbaden: Harrassowitz.
- Kobayashi, Masato. 2004. *Historical phonology of Old Indo-Aryan consonants*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Kümmel, Martin Joachim. 2012. Typology and reconstruction: The consonants and vowels of Proto-Indo-European. In Benedicte Nielsen Whitehead et al. (eds.), *The sound of Indo-European*, 292–329. Copenhagen: Museum Tusculanum.
- Kümmel, Martin Joachim. 2014. Syllable- and word-related developments in earlier Indo-Iranian. In Javier Caro Reina & Renata Szczepaniak (eds.), *Syllable and word languages*, 204–21. Berlin: de Gruyter.

- Kümmel, Martin Joachim. 2016a. Is ancient old and modern new? Fallacies of attestation and reconstruction (with special focus on Indo-Iranian). In David M. Goldstein, Stephanie W. Jamison & Brent Vine (eds.), *Proceedings of the 27th annual UCLA Indo-European conference*, 79–96. Bremen: Hempen.
- Kümmel, Martin Joachim. 2016b. *syá- im Indoiranischen: Zahlwort und Demonstrativum? In Andrew Miles Byrd, Jessica DeLisi & Mark Wenthe (eds.), *Tavet tat satyam: Studies in honor of Jared S. Klein on the occasion of his seventieth birthday*, 179–90. Ann Arbor, MI: Beech Stave.
- Kümmel, Martin Joachim. 2016c. Zur “Vokalisierung” der Laryngale im Indoiranischen. In Dieter Gunkel et al. (eds.), *Sahasram ati srajas: Indo-Iranian and Indo-European studies in honor of Stephanie W. Jamison*, 216–26. Ann Arbor, MI: Beech Stave.
- Kümmel, Martin Joachim. 2017. Agricultural terms in Indo-Iranian. In Martine Robbeets & Alexander Savelyev (eds.), *Language dispersal beyond farming*, 275–90. Amsterdam: Benjamins.
- Kümmel, Martin Joachim. 2018a. Anatolisches und indoiranisches Verbum: Erbe und Neuerung. In Elisabeth Rieken (ed.), *100 Jahre Entzifferung des Hethitischen: Morphosyntaktische Kategorien in Sprachgeschichte und Forschung*, 239–58. Wiesbaden: Reichert.
- Kümmel, Martin Joachim. 2018b. The morphology of Indo-Iranian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 1888–924. Berlin: De Gruyter Mouton.
- Kümmel, Martin Joachim. 2018c. The survival of laryngeals in Iranian. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 162–72. Ann Arbor, MI: Beech Stave.
- Kümmel, Martin Joachim. 2020. Substrata of Indo-Iranic and related questions. In Romain Garnier (ed.), *Loanwords and substrata*, 237–77. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Kuryłowicz, Jerzy. 1927. Les effets du *a* en indoiranien. *Prace filologiczne* 11. 201–43.
- Lipp, Reiner. 2009. *Die indogermanischen und einzelsprachlichen Palatale im Indoiranischen*. Vol. 1. *Neurekonstruktion, Nuristan-Sprachen, Genese der indoarischen Retroflexe, Indoarisch von Mitanni*; Vol. 2. *Thorn-Problem, indoiranische Laryngalkonsonanten*. Heidelberg: Winter.
- Lubotsky, Alexander. 2018. The phonology of Proto-Indo-Iranian. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 1875–88. Berlin: De Gruyter Mouton.
- Luján, Eugenio. 2019. On Indo-European superlative suffixes. *Indogermanische Forschungen* 124. 305–41.
- Malzahn, Melanie. 2010. *The Tocharian verbal system*. Leiden: Brill.
- Martirosyan, Hratch K. 2010. *Etymological dictionary of the Armenian inherited lexicon*. Leiden: Brill.
- Mayrhofer, Manfred. 1982. Welches Material aus dem Indo-Arischen von Mitanni verbleibt für eine selektive Darstellung? In Erich Neu (ed.), *Investigationes philologicae et comparativae: Gedenkschrift für Heinz Kronasser*, 72–90. Wiesbaden: Harrassowitz. Also in Rüdiger Schmitt (ed.), *Manfred Mayrhofer: Ausgewählte kleine Schriften*. Vol. 2, Wiesbaden: Reichert, 1996, 304–22.

- Mayrhofer, Manfred. 2004. Zur Vertretung der indogermanischen Liquiden in den indoiranischen Sprachen. *Indologica Taurinensia* 28. 149–61.
- Melchert, H. Craig. 1994. *Anatolian historical phonology*. Amsterdam: Rodopi.
- Melchert, H. Craig. 2012a. Luvo-Lycian dorsal stops revisited. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphophonemics*, 206–18. Munich: LINCOM Europa.
- Melchert, H. Craig. 2012b. Hittite *hi*-verbs of the type $\bar{a}C_1i$, aC_1C_1anzi . *Indogermanische Forschungen* 117. 173–85.
- Melchert, H. Craig. 2015. The Tocharian *s*-preterite. In Melanie Malzahn, Michaël Peyrot, Hannes Fellner & Theresa-Susanna Illés (eds.), *Tocharian texts in context*, 127–35. Bremen: Hempen.
- Melchert, H. Craig & Norbert Oettinger. 2009. Ablativ und Instrumental im Hethitischen und Indogermanischen: Ein Beitrag zur relativen Chronologie. *Incontri Linguistici* 32. 53–73.
- Miller, D. Gary. 1977a. Bartholomae's Law and an IE root constraint. In Paul J. Hopper (ed.), *Studies in descriptive and historical linguistics: Festschrift for Winfred P. Lehmann*, 365–92. Amsterdam: Benjamins.
- Miller, D. Gary. 1977b. Some theoretical and typological implications of an Indo-European root structure constraint. *Journal of Indo-European Studies* 5. 31–130.
- Olander, Thomas. 2019. Indo-European cladistic nomenclature. *Indogermanische Forschungen* 124. 231–44.
- Peters, Martin. 1991. 'Οπεσθ- neben' Οπεστ(ā)-. *Die Sprache* 35. 135–8.
- Peyrot, Michaël. 2018. Tocharian B *etswe* 'mule' and Eastern East Iranian. In Lucien van Beek et al. (eds.), *Farnah: Indo-Iranian and Indo-European studies in honor of Sasha Lubotsky*, 270–83. Ann Arbor, MI: Beech Stave.
- Pinault, Georges-Jean. 2006. Retour sur le numéral 'un' en Tokharien. *Indogermanische Forschungen* 111. 71–97.
- Pooth, Roland. 2011. Die 2. und 3. Person Dual und das Medium. In Thomas Krisch & Thomas Lindner (eds.), *Indogermanistik und Linguistik im Dialog*, 473–83. Wiesbaden: Reichert.
- Pyysalo, Jouna. 2013. *System PIE: The primary phoneme inventory and sound law system for Proto-Indo-European*. Helsinki: UnigrafiaOy. <http://ethesis.helsinki.fi>
- Rasmussen, Jens Elmegård. 1989. Die Tenues Aspiratae: Dreiteilung oder Vierteilung des indogermanischen Plosivsystems und die Konsequenzen dieser Frage für die Chronologie einer Glottalreihe. In Theo Vennemann (ed.), *The new sound of Indo-European: Essays in phonological reconstruction*, 153–76. Berlin: Mouton de Gruyter.
- Rau, Jeremy. 2009. *Indo-European nominal morphology: The decads and the Caland System*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Rieken, Elisabeth. 2010. Das Zeichen <sà> im Hieroglyphen-Luwischen. In Aygül Süel (ed.), *Acts of the VIIth international congress of Hittitology*. Vol. 2, 651–60. Ankara: T. C. Çorum Valiliği.
- Ringe, Don, Tandy Warnow & Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100. 59–129.
- Scharfe, Hartmut. 1996. Bartholomae's Law revisited or how the Ṛgveda is dialectically divided. *Studien zur Indologie und Iranistik* 20. 351–77.

- Schleicher, August. 1860. *Die Deutsche Sprache*. Stuttgart: Cotta.
- Schleicher, August. 1861, 1862. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. 2 vols. Weimar: Böhlau.
- Szemerényi, Oswald. 1976. The problem of Aryan loanwords in Anatolian. In *Scritti in onore di Giuliano Bonfante*. Vol. 2, 1063–70. Brescia: Paideia. Also in P. Considine & J. T. Hooker (eds.), *Scripta minora*. Vol. 4, 1671–8. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 1991.
- Tremblay, Xavier. 2005. Bildeten die iranischen Sprachen ursprünglich eine genetische Familie oder einen Sprachbund innerhalb des indo-iranischen Zweiges? Beiträge zur vergleichenden Grammatik der iranischen Sprachen V. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 673–88. Wiesbaden: Reichert.
- de Vaan, Michiel. 2009. The derivational history of Greek *ἵππος* and *ἵππεύς*. *Journal of Indo-European Studies* 37. 198–213.
- Viredaz, Rémy. 1983. Proto-Indo-European and North-West Caucasian vowel systems. Unpublished manuscript. <https://bit.ly/3PEkbbkZ> (accessed 24 May 2022).
- Weiss, Michael. 2018. Limited Latin Grassmann's Law: Do we need it? In Dieter Gunkel et al. (eds.), *Vina diem celebrent: Studies in linguistics, and philology in honor of Brent Vine*, 438–47. Ann Arbor, MI: Beech Stave.
- Werba, Chlodwig. 2005. Sanskrit *duhitár-* und ihre (indo-)iranischen Verwandten, Zur "Vokalisierung" der Laryngale im Ur(indo)arischen. In Günter Schweiger (ed.), *Indogermanica: Festschrift Gert Klingenschmitt*, 699–732. Tübingen: Schweiger VWT-Verlag.
- Werba, Chlodwig. 2016. Ur(indo)arischen im Nūristānī. Zur historischen Phonologie des Indoiranischen. In Andrew Miles Byrd, Jessica DeLisi & Mark Wenthe (eds.), *Tavet tat satyam: Studies in honor of Jared S. Klein on the occasion of his seventieth birthday*, 341–59. Ann Arbor, MI: Beech Stave.
- Woodhouse, Robert. 2012. Latin *radius* 'ray of light; spoke of wheel': PIE **a* versus an extension of *u*-umlaut in West Saxon. *Indogermanische Forschungen* 117. 1–13.
- Woodhouse, Robert. 2015. On the reality of the laryngeal theory: A response to Witold Mańczak. *Historische Sprachforschung* 126. 3–32.

Tijmen Pronk

15.1 Introduction

Since the times of Bopp and Schleicher, Baltic and Slavic have been treated as a single branch of the Indo-European language family. Throughout the nineteenth century, this view remained unchallenged, and it is presented as received wisdom in Brugmann's *Grundriss* (1897: 20–1). At the beginning of the twentieth century, however, Meillet (1905: 201–2; 1922: 40–8) challenged the idea of a Balto-Slavic unity and argued that those similarities between Baltic and Slavic that are not archaisms inherited from (dialectal) Proto-Indo-European are due to parallel innovations. Throughout the twentieth century, the matter remained controversial. Balto-Slavic unity was defended by Rozwadowski (1912) and Vaillant (1950: 14), for example, while scholars like Senn (1941; 1970), Fraenkel (1950: 73–112), Pohl (1992), Schmid (1992) and Andersen (1996) remained sceptical and explained the similarities in terms of language contact and convergence. During the last quarter of a century, the *communis opinio* appears to have moved firmly in favour of the idea that there was indeed a period of shared innovations between Baltic and Slavic directly following the disintegration of the Proto-Indo-European parent language. As Olander (2015: 24) aptly put it: “By tracing back the identical developments in the two branches to a common ancestor we obtain the simplest model of the relationship between Baltic and Slavic, without a notable loss of explanatory power”.

Recent overviews of the shared Baltic and Slavic features that are relevant for the Balto-Slavic question can be found in Hock (2004, 2005), Euler (2007: 10–15), Young (2017), Petit (2018) and Villanueva Svensson (in press). Excellent general overviews of the scholarly literature have been given by Petit (2004), Hock (2006) and Dini (2014: 200–13). This chapter will discuss the most compelling phonological, lexical and morphological evidence in favour of a Balto-Slavic clade, after which it will address dialectal variation within Proto-Balto-Slavic, the internal grouping of Balto-Slavic, external affiliations of Balto-Slavic and linguistic contacts of Proto-Balto-Slavic.

First, however, it is useful to take a brief look at Balto-Slavic from an archaeological and palaeogenetic perspective.

Anthony (2007: 348) associated Balto-Slavic (pre-Baltic and pre-Slavic) with the Middle Dnieper culture that lasted from approximately 2800–2600 until 1900–1800 BCE. This is consistent with the linguistic evidence that the speakers of Balto-Slavic practised little agriculture (Pronk & Pronk-Tiethoff 2018: 304–8). Together with the closely related Fatyanovo culture to its north-east, the Middle Dnieper culture covers the area in which Baltic- or Balto-Slavic-looking hydronyms are found (Gimbutas 1963: 91; Anthony 2007: 380). Both these cultures belong to the larger Corded Ware horizon.

The split between Baltic and Slavic must have taken place a long time after the split of Balto-Slavic from other Indo-European groups in view of the large number of Balto-Slavic innovations. A date much before the beginning of the second millennium BCE is therefore unlikely. This makes it questionable whether the people who introduced genes from the Pontic-Caspian Steppe into the Baltic region during the third millennium BCE (Mittnik et al. 2018) and the people of the Rzucewo or Bay Coast Culture of the same period were speakers of early Baltic (pace Rimantienė 1992). They might have been the ancestors of Balto-Slavic speakers, as suggested by Kortlandt (2018a), in which case the idea that Balto-Slavic was still spoken on the Middle Dnieper during the third millennium BCE must be rejected. It seems more likely that the people who brought steppe genes into the Baltic region in the third millennium spoke another, now lost, dialect of Indo-European (cf. Kortlandt 2018a).

In the basin of the Dnieper river, the speakers of Balto-Slavic apparently picked up names for fish such as the wels catfish (Lith. *šāmas*, Ru. *som*), tench (Lith. *lŷnas*, Ru. *lin'*), sturgeon (OPr. *esketres*, Ru. *osĕtr*) and perhaps ruffe (Lith. *ež(e)gŷs*, Pol. *jażdż*, *jazgarz*).¹ The importance of rivers and fishing for the speakers of Balto-Slavic may also be reflected in the fact that Baltic and Slavic uniquely share verbs for wading (Lith. 3pres. *breñda*, Ru. 1sg.pres. *bredŭ*) and diving (Lith. *nĕrti*, RuCS *vb-nrĕti*), and nouns for spawning (Lith. *neřštas*, Ru. *nĕrest*), dugout canoe (Lith. *eldijà*, OCS *aldii*) and raft (Latv. *pluts*, Ru. *plot*). The Baltic name for the pike (Lith. *lydŷs*, OPr. *liede*), a fish that was an important food source in the Baltic area during the Neolithic (Rimantienė 1992: 105), has no cognate in Slavic, but this could be due to a later replacement.

From the middle Dnieper region, the ancestors of the speakers of West and East Baltic would have moved along the rivers into the forests to the north, where they borrowed words for woodland animals such as the elk (Lith. *briedis*, Latv. *briēdis*, OPr. *braydis*), woodpecker (Lith. *genŷs*, Latv. *dzenis*,

¹ Because of the different vowels in the suffix, it seems likely that Lith. *lašišà* and Ru. *losos'* 'salmon' were borrowed independently from similar sources, as was OHG *lahs* 'salmon'.

OPr. *genix*), hawk (Lith. *vānagas*, Latv. *vanags*, OPr. *spergla-wanag* ‘sparrow-hawk’) and perhaps bear (Lith. *lokys*, Latv. *lācis*, OPr. *clokis*) from an unknown non-Indo-European language. Because there are very few shared innovations between Old Prussian and East Baltic (see [Section 15.3.2](#)), it would seem likely that they were spoken by different groups shortly after the migrations to the north and north-west from the Dnieper basin. Most if not all common East Baltic innovations, including the creation of new locative cases due to contact with another, most probably Uralic language, could have taken place before the East Baltic languages entered the Baltic coastal areas.

The speakers of pre-Proto-Slavic would originally have occupied the area between the Middle Dnieper and Upper Dniester ([Anthony 2007](#): 379–80). Before their spread across Central and Eastern Europe after 500 CE, they can be most probably located to the north-east of the Carpathian mountains ([Udolph 1979](#): 619–23) and have often been associated with the Zarubintsy culture (appr. 300 BCE–100 CE, see e.g. Maksimov in [Rusanova & Symonovič 1993](#): 36–9).

A study of the Y chromosome of Slavic populations supports the hypothesis that the Slavic expansion started from present-day Ukraine ([Rębała et al. 2007](#)). So far, no support for Proto-Balto-Slavic has been found in studies of DNA. [Rębała et al. \(2007\)](#) found significant differences in Y-chromosomal haplogroup distribution between Slavic and Baltic populations. Baltic populations are genetically the closest to East Slavs, but this is probably due to a Baltic substrate in northern East Slavic ([Kushniarevich et al. 2015](#)).

15.2 Evidence for the Balto-Slavic Branch

15.2.1 Phonology and Relative Chronology

In a 2005 article, [Matasović \(2005b\)](#) discussed the following eleven phonological innovations that are found in Baltic and Slavic:

1. depalatalizations of palatovelars
2. satemization
3. the *ruki* rule
4. Hirt’s Law²
5. the development of syllabic resonants
6. Lidén’s Law³
7. loss of word-final **-d*
8. Winter’s Law⁴
9. **o > *a*

² I.e. a stress retraction onto a preceding syllable in which the nucleus was followed by a laryngeal.

³ I.e. loss of word-initial **u-* before **-r-*, perhaps also before **-l-*.

⁴ I.e. lengthening of a preceding vowel and introduction of acute intonation in a preceding syllable by what are traditionally reconstructed as voiced unaspirated stops.

10. deaspiration of the aspirated stops⁵

11. loss of laryngeals.⁶

Matasović concluded that these innovations could have occurred in the same chronological order and that no Baltic or Slavic innovation can be shown to have occurred before these innovations. The relative chronology of Balto-Slavic sound changes set up by Kortlandt (2011: 157–76; 2009: 43–6) leads to the same conclusion. The list of shared innovations can be extended by adding, e.g., the evolution of Baltic and Slavic mobile accentuation (Pedersen 1933; Olander 2009, 2019; Jasanoff 2017; Kortlandt 2018b). The exact phonetic conditions of some of the sound laws and their exact chronological order remain a matter of debate (cf. Hock 2006 with ample references to the relevant literature), but this does not affect the conclusion that Baltic and Slavic had a long shared history after Proto-Indo-European had dissolved.

15.2.2 *Shared Innovations in the Core Lexicon*

The existence of a unitary Balto-Slavic proto-language is confirmed by the fact that Baltic and Slavic share a number of lexemes belonging to the core vocabulary that are either not found in other Indo-European languages or that show identical morphological or semantic innovations compared to cognates in other Indo-European languages. The examples can easily be drawn from Trautmann's 1923 dictionary or from Ślawski 1970. The following seventeen etyma with a meaning that is usually thought to belong to the core vocabulary are exclusively Balto-Slavic: *put- 'bird', *konzd- 'to bite', *skeit- 'to count', *touzk- 'fat', *nog- 'foot, leg', *ronkaz 'hand, arm', *golzua? 'head', *rogos 'horn', *ledus 'ice', *ke/ol- 'knee', *edzero 'lake', *uelk- 'to pull', *dz/guaizd- 'star', *solzdus 'sweet', *met- 'to throw', *bo/ēlz- 'white', *su(n) 'with'. Based on the 1971 Swadesh 100 list or the 2019 Jena 170 list (see www.eva.mpg.de/linguistic-and-cultural-evolution/research/ie-cor/) of core lexical meanings, this amounts to around 10 per cent of the total reconstructed Proto-Balto-Slavic basic lexicon.

15.2.3 *Shared Morphological Innovations*

There are numerous shared innovations between Baltic and Slavic in morphology. The following list is far from complete, but it contains those items that

⁵ I.e. merger of what are traditionally reconstructed as mediae and mediae aspiratae.

⁶ This should be changed into the merger of the laryngeals into a single segment, probably a glottal stop. The eventual loss of this segment occurred independently in Baltic and Slavic in view of OCS *kamy* 'stone' < *kaHmōn, with metathesis from PIE *h₂ekmōn, but Lith. *akmuō* 'stone' without metathesis (although Matasović 2005b: 152 does not consider this evidence to be conclusive). On the dating of the loss of the laryngeals as segments in Balto-Slavic, see also Kortlandt 2009: 6.

are fairly indisputable. For these and other proposed shared innovations, the reader is referred to the literature cited in the introduction, especially [Hock 2005](#) and [Villanueva Svensson in press](#), as well as [Stang 1966](#): 18–20, [Gołab 1992](#): 50–1, and [Kortlandt 2016c](#), 2018c.

Shared innovated nominal endings:

- *o*-stem gen.sg. **-ā* (Lith. *-o*, OCS *-a*, in OPr. *-as* enlarged with *-s*, see below) < PIE abl. **-oed*
- the generalized consonant stem gen.sg. **-es* (OLith., OPr. *-es*, OCS *-e*) ← PIE **-es*, **-os*
- consonant stem instr.pl. **-mizs* (Lith. *-mis*, OCS *-mi*) ← PIE **-b^his*(?)
- adjectival *o*-stem neuter nom.acc.sg. **-o* (Lith. *-a*, OPr. *-a*, OCS *-o*) < PIE pronominal **-od*

Shared innovations in nominal derivation:

- deadjectival abstracts and nomina actionis in **-b-* (Lith. *-ba*, *-yba*, *-ybė*, OCS *-ьба*, *зѣлоба*, *зѣлоба* ‘malice’, [Arumaa 1955](#); probably from PIE **-b^hh₂-* ‘to become’)
- deverbal abstracts in **-imo* (Lith. *-imas*, OCS *-ьмо*, ultimately < PIE **-mn-* ([Pronk 2014](#)))
- grammaticalization of the adverbial ending **-ai* (Lith., OPr. *-ai*, OCS *-ě*) < PIE loc.sg. **-oi*

Shared innovations in the morphology of the verbal system:

- preterits/aorists in **-ā* (Lith. *-o*, OPr. *-a*, OCS aor. *-a*)
- verbs with pres. **-ou̯je/o-*, pret. **-ou̯ā* (Lith. pres. *-auja*, pret. *-avo*, OPr. 3pres. *-awie*, OCS pres. *-uję*, aor. *-ova*)
- statives in **-eɣ-* with an *i*-present (OPr. *turīt*, *turri* ‘have’, Lith. *budėti*, *būdi*, ORu. *būdėti*, *būdim̃* ‘be awake’)
- perfects joining the preceding category (Lith. *garėti*, *gāri* ‘evaporate’, ORu. *gorėti*, *gorit̃* ‘burn’)
- transformation to a thematic present of PIE perf. **mog^h-* ‘be able’
- present stems **doɣd-* ‘give’ and **ded-* ‘put’ (OLith. *duosti*, *dest*, OPr. *dāst*, OCS *dastь*, *-deždę*) ← PIE pres. **di/e-deh₃-*, **d^hi/e-d^heh₁-*
- 2sg. pres. **esezi* ‘you are’ ← PIE **h₁esi* (Lith. *esi*, OPr. *assai*, OCS *jesi*) ([Kortlandt 2009](#): 156)
- causatives in **(e)i-* (Lith. *báudinti*, *báudyti* ‘urge’, OCS *vbz-buditi* ‘awaken’) ← PIE **-eie-*
- oblique forms of the masculine and neuter present active participle in **-ont-je/o-* (e.g., gen.sg.m. Lith. *nėsančio*, OCS *nesqšta* ‘carrying’)
- infinitives in **-izt(e)i* with analogical **-ɣ-* after infinitives in **-ert(e)i* and **-azt(e)i* (Lith. *-yti*, *-ėti*, *-oti*, OCS *-iti*, *-ėti*, *-ati*)

Further, there are some nouns in which Baltic and Slavic have (near) identical derivatives from Indo-European roots. In [Trautmann’s 1923](#) dictionary we find, inter alia, Lith. *āvinas*, OPr. *awins*, ORu. *овѣнъ* ‘ram’, Lith. *artójas*, OPr.

artoys, OCz. *rataj* ‘ploughman’, Lith. *plaūčiai*, OPr. *plauti*, OCS *plūšta* ‘lung-
(s)’, Lith. dial. *péntis*, OPr. *pentis*, OCS *pęta* ‘heel’.

15.2.4 *Shared Syntactic Innovations*

Due to the difficulty of reconstructing Proto-Indo-European syntax, it is also difficult to identify any syntactic innovations that Baltic and Slavic may have shared. In general, there are few methodological tools that we can use to determine whether any similarities in the structural properties of Baltic and Slavic are due to shared inheritance, shared innovation, independent innovation or mutual influence. Therefore, “the issue of Balto-Slavic ‘unity’ . . . should center around phonology, morphology, and the lexicon” (Holvoet 2018: 2001).

A seemingly shared Balto-Slavic syntactic feature is reflected in the definite adjectives that are attested in both branches, e.g. Lith. *geràsis*, OCS *dobryi* ‘good’. These definite adjectives derive from a nominal sentence in which a relative pronoun connects two nominal forms, agreeing in case, number and gender with the first of these nominals (Petit 2009). Parallels for such a construction are found in Iranian (Meillet 1922: 44). This syntactic construction “predat[es] at least the split between Balto-Slavic and Indo-Iranian” (Widmer et al. 2017: 811) and is likely to be an archaism inherited from PIE (Petit 2009: 354–5). The only plausible shared Balto-Slavic syntactic innovation reflected in the definite adjectives is the agreement between the relative pronoun and the head of the construction, which is also found in Iranian (Petit 2009: 354–5).

The most promising example of a syntactic innovation that is shared by Baltic and Slavic only and less likely to have arisen independently or as a result of contact between Baltic and Slavic is the complete loss of the Proto-Indo-European middle voice and its replacement by reflexive verbs in at least some of its functions. See Holvoet 2020 for an extensive discussion of this issue.

15.3 **The Internal Structure of Balto-Slavic**

15.3.1 *Proto-Balto-Slavic Dialectal Differentiation*

One might wonder whether any dialectal differentiation that might have been present in Proto-Balto-Slavic was carried over into Baltic and Slavic. According to Olander (2015: 24) “there are cases of variation that cannot be avoided in a reconstructed Balto-Slavic proto-language, such as the existence of different lexemes for the same notion, or the existence of variants with initial **a* or **e* in the same lexeme in different areas (Andersen 1996: 206 and passim)”. Because the lexical data is open to various interpretations, I will here focus on the variants with initial **a* or **e*, such as Ru. *orël* but Lith. *erėlis* ‘eagle’ < PIE **h₂er-l-*.

Andersen proposed a scenario in which the variation arose within a Baltic-Slavic dialect continuum, even before some of the common Balto-Slavic innovations mentioned at the beginning of this chapter (1996: 106–7). The dialectal variants would have continued to coexist throughout the Proto-Slavic and Proto-East-Baltic periods and, in some cases, in the modern Slavic and Baltic languages. Such a long period of coexisting variants of the same words is highly unlikely and not supported by the data. Instead, branch-internal mechanisms caused the rise of the variation in initial vocalism.

In Slavic, it has long been clear that the variation between initial *je-* (< **e-*, **je-* or **ja-*) and *o-* (< **a-*) cannot be separated from that between *u-* and *ju-* in OCS *utro*, *jutro*, or that between *a-* and *ja-* in OCS *aviti*, *javiti*, ORu. *azъ*, *jazъ*. The variation is due to sandhi variants that arose when a yod developed in hiatus between two vowels, one of which was a front vowel (Pedersen 1905: 311). Similarly, words with an initial vowel developed a sandhi variant with initial **u-* if they were preceded by a word ending in a rounded vowel, e.g. Cz. *vejce* ‘egg’ < **ajbce*. Some instances of initial *je-* are the result of the regular umlaut **ja-* > **jä-* > **je-* and thus originally positional variants of **a-* > **o-*. The alternations between initial **uo-*, **je-* and **o-* and between **e-* and **je-* in sandhi led to the generalization of one of the variants, and sometimes to the analogical introduction of an etymologically “incorrect” onset, e.g. in the word for ‘wasp’, which is **osa* in almost all of Slavic, but *vosa* in Czech. The Czech form is the older variant in view of outer-Slavic cognates such as Lith. *vapsvā* and Lat. *vespa*. The variant **osa* must be due to reinterpretation of **vosa* as a sandhi variant after rounded vowels (Pedersen 1905: 312).

There is no reason to assume that the Baltic variation between initial *a-* and *e-* and the Slavic alternation between initial *o-* and *je-* are in any way related (see further Derksen 2002; Kortlandt 2011: 255–8). They therefore provide no evidence for a Balto-Slavic dialect continuum, nor for a shared innovation.

The strongest potential evidence for inner-Balto-Slavic variation that I am aware of is the 1sg. personal pronoun **h₁eg̃*, that underwent Winter’s Law (> Proto-Balto-Slavic **eʔdʒ*) and produced ORu. *ja*. In Baltic, the same pronoun has a voiceless sibilant and a short vowel: OLith. *eš*, Latv. *es*, OPr. *as*, *es*. The Baltic forms seem to suggest that there was a positional variant **h₁ek̃* before a following word beginning with a voiceless consonant that did not undergo Winter’s Law. If this is correct, Slavic and Baltic may have generalized different sandhi variants. The generalization of one of the variants could of course have happened at any point after Winter’s Law, and not necessarily before the dissolution of Proto-Balto-Slavic. Other explanations are also conceivable. Kortlandt (2013a), for example, argued that the Baltic forms and Slavic **ja* are the result of post-Proto-Balto-Slavic shortenings of original **eʔdʒum*, preserved in Slavic as *(*j*)*azъ* (e.g. ORu. *jazъ*). In either

scenario, there is no compelling evidence for internal differentiation within Proto-Balto-Slavic that was carried over into Baltic or Slavic.

15.3.2 *Internal Grouping*

Traditionally, Balto-Slavic has been divided into Baltic and Slavic, with a further split between West and East Baltic after a period of common Baltic innovations. The separate status of Slavic is evident, but the existence of a period of common Baltic innovations is more difficult to demonstrate; see most recently Villanueva Svensson 2014, Hill 2016 and Kortlandt 2018c with references to the older literature. Stang (1966: 2–10) lists the similarities between the Baltic languages that set them apart from all other Indo-European languages, including Slavic (notation as in the original):

- complete merger of the 3sg. and 3pl. verbal endings
- two preterit classes in **-ē* and **-ā*
- a distribution between the 3rd person verbal endings **-ti* to monosyllabic stems and **-t > zero* to polysyllabic stems
- 1sg. athematic **-māi*
- a thematic vowel *-a-* < **-o-*, never **-e-*
- nominal *ē*-stems
- intrusive **k* before consonant clusters beginning with **s*
- nomina actionis with the suffix **-sĭan-*, perhaps also **-sen-*
- nouns in **-ūnas*
- diminutive suffixes **-ēlija-*, **-už-*, **-ut-*, **-ait-* (also in patronymics)
- adjectives in **-ing-*
- identical compound names, often with a binding vowel **-i-*
- *ā*-presents to verbs in **-īti*
- *sta*-presents to middle/intransitive verbs
- causatives in **-ina-*
- a large amount of uniquely shared lexicon, including identical derivatives from inherited roots and semantic innovations in inherited material (cf. Petit 2010: 10–11).

To these we can add the loss of **-j-* between a consonant and a front vowel (Villanueva Svensson 2014: 165) and the identical restructuring of some Proto-Indo-European consonant stems and root nouns: Lith. *akis*, OPr. *ackis* ‘eye’, Lith. *ausis*, OPr. acc.pl. *āusins* ‘ear’ (Hill 2016: 210–11), Lith. *saulė*, OPr. *saule* ‘sun’, Lith. *gėrvė*, OPr. *gerwe* ‘crane’, Lith. *žemė*, OPr. *semmē* ‘earth’, Lith. *dienà*, OPr. acc.sg. *deinan* ‘day’. Other proposed shared innovations, such as the change of **-ijā* to **-ē* (Petit 2010: 6; Villanueva Svensson 2014: 165; cf. also Hill 2016) and the shortening of unstressed **-ī* < **-eie-* (Hill 2016: 214–22; Villanueva Svensson 2019), remain the subject of debate.

In the former case, if there was a raising of **-ijā* to **-ijē*, it may well have been shared by Slavic, cf. the type OCS *ml̋nii* (f.) ‘lightning’ < **-ijē*. This leaves the contraction and associated metatony as potentially shared Baltic innovations, but consider the general preservation of **ā* after yod in other positions (e.g. Lith. *jóti* ‘to ride’, *bijóti* ‘to fear’, *valià* ‘will’ etc.) and further objections raised by Kortlandt (2018c). The alleged change of **-ijā* to **-ē* thus remains poorly understood and cannot serve as evidence for the branching of Balto-Slavic.

Most evidence for Hill’s contraction of unstressed **-ī-* < **-eie-* is judged to be inconclusive by Villanueva Svensson (2019), except for the PIE *i*-stem dat. sg. ending **-eiei*, for which the common Baltic evidence would be the *ti*-stem dative **-ti* < **-teiei* (Skt. *-taye*) that was grammaticalized as an infinitive (Lith. *-ti*, Latv., OPr. *-t*). We are thus dealing with a sound law that explains only a single morpheme, which weakens it considerably. Moreover, the Baltic infinitive ending **-ti* has a potential counterpart in Slavic. Next to the well-known Slavic infinitive ending **-ti*, there is a widespread variant **-tb*, which could go back to Balto-Slavic **-ti*. There cannot have been a general reduction of unstressed **-i* to **-b* in Slavic, because nominal endings in *-i*, e.g. several forms of the *i*-stems, nom.pl. *-i* in the *o*-stems, instr.pl. *-mi* etc. are never reduced (cf. Vaillant 1950: 219–20). This means that the shortening in the infinitive of unstressed **-tī* > **-ti* > **-tb*, if that is indeed how the Slavic variants arose, only affected the specific pre-Proto-Slavic sequence that produced *-i* in the infinitive and perhaps in the athematic imperative, cf. OCS *daždь* ‘give!’ < 2sg. optative **-ieh₁-s(?)*. However, it did not affect the dat.sg. ending *-i* of the *i*- and *u*-stems, which was also unstressed. In short: the Baltic infinitive ending **-ti* has a potential parallel in Slavic, so the alleged shortening of an alleged Proto-Balto-Slavic infinitive ending **-tī* cannot be used as evidence for a Proto-Baltic stage.

Many of the shared features of West and East Baltic can be and have been argued to be either inherited from Proto-Balto-Slavic and lost in Slavic or independent innovations, most prominently by Kortlandt (2018c with references to earlier works). In order to demonstrate that there was indeed a period of shared Baltic innovations, the innovated feature must not only be shared by West and East Baltic, it must also be shown to have never existed in Slavic, and its introduction should not be a trivial development. Few of the shared features collected by Stang and others fulfil these criteria. The shared derivational suffixes on Stang’s list could all have been lost in Slavic. The same is true for lexical items such as Lith. *turėti*, Latv. *turēt*, OPr. *turrītwei* ‘to have’ and Lith. *gimti*, Latv. *dzimt* ‘to be born’, OPr. *gemmons* ‘born’. The semantic innovation in Lith. *girià*, Latv. *dziņa* ‘forest’, OPr. *garian* ‘tree’ versus OCS *gora* ‘mountain’ turns out to be trivial if one takes a closer look at the semantics of the Slavic cognates, cf. Bulg. *gora*

and Slk. *hora* ‘forest’. The word appears to have designated a wooded slope or mountain in Proto-Slavic and Proto-Balto-Slavic. The semantic innovation in Lith. *mėdis* ‘tree’, Latv. *mežs*, OPr. *median* ‘forest’ versus OCS *mežda* ‘boundary’ is also trivial, cf. Sln. dial. *mej* ‘forest’ from the same root and the connection between Lith. *vidūs* ‘middle’ and Old English *widu* ‘wood’.

The most robust evidence for a Proto-Baltic period is, in my view, presented by the productivity of nominal *ē*-stems (whatever their origin), the (near) merger of 3sg. and 3pl. verbal forms, the loss of **j-* between a consonant and a front vowel and the identical evolution of a number of former consonant stems and root nouns. This seems to suggest that there was indeed a Proto-Baltic period, which lasted for at least a few generations but probably no longer than a few centuries.

It has long been clear that West and East Baltic are also separated by some isoglosses that connect East Baltic with Slavic. The most often cited examples are the following (see Villanueva Svensson *in press* for a few more inconclusive examples):

- the *o*-stem gen.sg. ending (Lith. *-o*, OCS *-a* < PIE abl.sg. **-oed* versus OPr. *-as*)
- the initial consonant in the word for ‘nine’ (Lith. *devyni*, OCS *devętъ* versus OPr. *newīnts* ‘ninth’)
- the word for ‘third’ (Lith. *trėčias*, OCS *tretii* versus OPr. *tīrts*, *tirtis*)
- presence versus absence of *-s-* in the dat.sg. and loc.sg. of the demonstrative pronoun (Lith. *tāmui*, *tamė*, *tai*, *tojė*, OCS *tomu*, *tombъ*, *toi* versus OPr. *stesmu*, *stessei*).

It is, however, uncertain that these isoglosses are the result of shared innovations of only East Baltic and Slavic. In the first three cases, East Baltic and Slavic may preserve the Proto-Balto-Slavic situation, and in the fourth case they may have innovated independently.

The Prussian *o*-stem gen.sg. ending *-as* has been explained from PIE **-oso*, **-osio*, **-os*, as analogical to the feminine *ā*-stem ending *-as* (Leskien 1876: 31–3), or from the same **-oed* as East Baltic with addition of the genitive singular marker **-s* (Vaillant 1958: 30; see further Rinkevičius 2015: 106–7 with literature). The latter explanation seems to be the least problematic phonetically, and it has been suggested that traces of an earlier *s*-less ending *-a*, *-u* may exist within Old Prussian (Leskien 1876: 33–4; Girdenis & Rosinas 1977: 3; Kortlandt 2009: 192). There is therefore no demonstrably old distinction between West and East Baltic in this ending.

The introduction of *d-* in ‘nine’ (see above) is due to anticipation of the *d-* of ‘ten’ when counting. It is plausible that it first affected the cardinal and then spread to the ordinal numbers. For Proto-Balto-Slavic, one may then

reconstruct **deuin* ‘nine’, **neūintas* ‘ninth’, with preservation of the latter in OPr. *newīnts*.⁷

It is possible that East Baltic and Slavic shared the replacement of **tirtiōs* ‘third’, reflected in OPr. *tīrts*, *tirtis*, by **tretiōs*. It is, however, equally conceivable that the Prussian word was influenced by **ketūirtas* ‘fourth’ after the dissolution of Proto-Balto-Slavic (Mažiulis 2013: 912). It would then replace earlier **tretiōs*, which is itself best understood as a replacement of an even older **tritiōs*, cf. Av. *θritiia-*, Lat. *tertius*, Goth. *þridja* < **tri-t(H)-iHo-*, on the basis of **trejes* ‘three’. If that is the case, the resemblance between OPr. *tirtis* and Skt. *trītya-* ‘third’ is coincidental.

The analogical removal of *-s-* in the pronominal dat.sg. and loc.sg. Lith. *tāmui*, *tamè*, *tái*, *tojè* and OCS *tomu*, *tomь*, *toi* was an innovation in contrast to its preservation in OPr. *stesmu*, *stessei* ‘that’, cf. Skt. *tásmāi*, *tásmin*, *tásyai*, *tásyām*.⁸ The replacement was part of the general loss of the distinction between the direct and oblique cases in the pronoun, cf. OPr. dat.sg.f. *tennei* ‘her’ ← **tenness(i)ei* after nom.sg. *tennā* ‘she’, but preservation of dat.sg.f. *stessiei* to nom.sg.f. *stai*. It is conceivable that the removal of *-s-* occurred independently in East Baltic and Slavic, as in OPr. *tennei*. The removal of *-s-* was ultimately the result of the elimination of the suppletive nominatives m. **sa* and f. **saH*, which probably took place after the dissolution of Proto-Balto-Slavic as well (Kortlandt 2009: 139).

It seems most likely that, after the dissolution of Proto-Balto-Slavic, West and East Baltic remained a single unit for a relatively short period. There may have been a few shared innovations between East Baltic and Slavic during this same period, although the evidence is not very robust. If this is indeed the case, however, the dissolution of Balto-Slavic could be seen as a gradual process with increasing dialectal differences, “with East Baltic as an intermediate dialect between West Baltic and Slavic” (Kortlandt 2018c: 176).

15.4 The Relationship of Balto-Slavic to the Other Branches

15.4.1 Genealogical Relations

The perpetual question as to whether there was a period of shared Balto-Slavic and Germanic innovations is probably to be answered in the negative. The key argument has always been the **-m-* of the dat. and instr.du.pl. endings in Balto-Slavic (pl. OLith. *-mus*, *-mis*, OCS *-mь*, *-mi*) and the dat.pl. in Germanic (Goth.,

⁷ It cannot be ruled out either that the *n-* of Old Prussian is due to German influence (Derksen 2015: 126).

⁸ Hill’s (2016: 224–7) explanation of the loss of **-s-* as phonetic in unstressed position before **-m-* is unconvincing. This highly specific and phonetically problematic sound law is set up to explain the single morpheme **tōsm-*. It does not account for the feminine forms.

OHG *-m*) that contrast with **-b^h-* in the instr.pl. in Greek (*-φι*) and Armenian (*-b*), dat. and instr.du.pl. in Indo-Iranian (pl. Skt. *-bhyas*, *-bhis*) and dat.pl. in Italo-Celtic (Lat. *-bus*, OIr. *-b*). Because **-b^h-* is most clearly at home in the PIE instrumental plural ending, and **-m-* cannot have arisen out of thin air, it is likely that the Germanic and Balto-Slavic dative plural endings are archaic (Hirt 1895; Beekes 2011: 188). In other words, the Core Indo-European ending contained an **-m-*, which was replaced by **-b^h-* from the instrumental in Latin and Indo-Iranian, while in Slavic instrumental **-b^h-* itself was replaced by **-m-* from the dative (see Olander 2015: 269–70 for alternative views). It is clear that a common innovation of the dat.pl. ending in Germanic and Balto-Slavic cannot be substantiated. There are no other common innovations in the nominal declension (Leskien 1876), nor are there any shared phonological innovations. Parallel syntactic structures, such as the absolute dative or the genitive of negation, cannot be used as evidence because they can represent (partial) archaisms or reflect parallel innovations. Any evidence for a period of shared Germano-Balto-Slavic innovations must thus come from the lexicon, nominal derivation or verbal inflection.

A significant part of the vocabulary that is shared exclusively by Germanic and Balto-Slavic, collected and discussed by Stang (1972) and Nepokupnyj et al. (1989), consists of words belonging to semantic fields that are prone to borrowing, such as flora and fauna. Some of the correspondences from semantic fields other than flora and fauna could easily be archaisms inherited from Proto-Indo-European, e.g. Goth. *ju*, Lith. *jaiū*, OCS (*j*)*u-že* ‘already’ < PIE **h₂iew*; ON *lyðr*, Lith. *liáudis*, OCS *ľudije* ‘people’ < PIE **h₁leud^h-i-*; ON *ljóðr*, OCz. *l’ud* ‘people’ < PIE **h₁leud^h-o-*; MLG *noster(en)* ‘nostril’, Lith. *nasrai* ‘snout’, OCS *nozdrī* ‘nostrils’ < PIE **nh₂-(e)s-r-*; ON *súrr* ‘sour, bitter’, Latv. *sūrs* ‘salty, bitter’, OCS *syrъ* ‘damp’ < PIE **suH-ro-*; OPr. *tūsīmtons*, OCS *tysęšti*, Got. *þusundi* ‘thousand’ < PIE **tuHs-dkmt-*. The remaining shared vocabulary does not contain any obvious replacements of Proto-Indo-European basic vocabulary and is not numerous enough to warrant the reconstruction of a period of joint Germanic and Balto-Slavic innovations.

A morphological argument often adduced in favour of a Germano-Balto-Slavic node is the shared adjectival suffix **-isko-*, Goth. *-isks*, Lith. *-iškas*, OCS *-bъskъ*, which primarily indicates origin from a particular place (Kluge 1926: 104). The suffix may have been created by adding adjectival **-ko-* to local adverbs in **-is* of the type Skt. *bahīh* ‘outside’, *āvíh* ‘manifestly’. If there was no Germano-Balto-Slavic node, the suffix must have arisen in a small number of forms in Proto-Indo-European and have become productive independently in Germanic and Balto-Slavic but have been lost elsewhere. This is conceivable. Vaillant’s (1958: 682) idea that the Slavic suffix was borrowed from Germanic and the Baltic one from Slavic seems unlikely, especially in view of Lithuanian *-š-*.

Another innovation perhaps shared between Balto-Slavic and Germanic is found in the semantics of nasal presents (Villanueva-Svensson 2011 with references). It has long been recognized that nasal presents in these languages are predominantly intransitive and have inchoative or fientive semantics, e.g. Goth. *ga-waknan* ‘to wake up’, Lith. *už-mìgti*, *-miñga* ‘to fall asleep’, OCS *vbz-bnqti* ‘to wake up’. In other branches, nasal presents typically form causatives, factitives and intensives (see Meiser 1993 with references), but cf. Lat. *-cumbō* ‘lie down’. In Greek, Indo-Iranian, Tocharian and Anatolian, nasal presents are mostly transitive in the active form, though not exclusively, cf., e.g., Gr. *φθίω* ‘to decline, decay’. Some nasal presents in Balto-Slavic, on the other hand, are transitive, e.g. Lith. *gáuna* ‘to obtain’ and OCS *tbknq* ‘to stab’.

The question as to whether the semantics of those Germanic and Balto-Slavic nasal presents that are inchoatives or fientives reflect a shared innovation depends on the reconstruction of the (pre-)Proto-Indo-European function of the nasal verbal suffix. Old Indo-European nasal presents are typically formed to roots with telic semantics. The nasal present appears to signify change of state (rather than “starkes Betroffensein”, Meiser 1993: 295) of the object of a transitive verb (cf. PIE **ui-n-d-* ‘find’) or the subject of an intransitive (unaccusative) verb. In addition, it is relevant that the suffix became a present marker and is never found in the aorist or perfect. This means that the oldest layer of nasal presents must have had progressive or ingressive semantics. They would thus have described the process of a change of state of either subject or object. Whether the nasal presents ended up as factitives and causatives or inchoatives and fientives depended on whether they were derived from a transitive or intransitive base. It has been argued that the intransitive Germanic and Balto-Slavic nasal presents derive from intransitive thematic aorists (Stang 1966: 340 for Balto-Slavic), middle root aorists (Kortlandt 2010: 219–20 for Germanic) or from the middle of the nasal present (Villanueva-Svensson 2011: 43; Kroonen 2012: 270 n. 11 for Germanic; cf. also Meiser 1993: 291–3). At least in Baltic, some nasal presents were derived from perfects: Lith. *kañka* ‘hang’, *rañda* ‘find’, *tañpa* ‘become’, *prañta* ‘acquire a habit or inclination’ (Stang 1966: 313, 315). The productivity of transitive or intransitive nasal presents, or indeed the lack of them, could be taken as a potential shared innovation of some branches of Indo-European, but it is a rather trivial development as long as it is assumed that both types existed in Proto-Indo-European. As an argument for a Balto-Slavo-Germanic node, the semantics of the nasal present are not particularly forceful.

A closer relationship between Balto-Slavic and any of the other branches is difficult to demonstrate as well. According to Kortlandt (2016a), “[t]he closest relatives of Balto-Slavic are Albanian and Indo-Iranian”, but shared innovations are few. Potentially shared phonological innovations are satemization,

which is also shared with Armenian, and the *ruki* rule, which possibly affected Armenian as well. In both cases, the shared innovation would have been the initial phonetic development, because the phonemicization of the rules is branch specific. Because phonetic changes can be reversed, it is impossible to show that none of the other branches took part in the initial, phonetic stages of satemization or the *ruki* rule as well. Consider in this respect the alleged *satem* reflexes in Luwic (Melchert 2012 with literature) and the Hieroglyphic Luwian sign *sa*₃, which occurs mainly in the vicinity of the *ruki* sounds (Rieken 2010).

Kortlandt (2018d: 287) proposed that the loss of a laryngeal between two vowels was a shared innovation of Balto-Slavic and Indo-Iranian. Laryngeals were also lost in this position in all other branches of Indo-European except Anatolian. In Greek, this loss produced a disyllabic sequence, but in Indo-Iranian and Balto-Slavic the result is a monosyllabic long vowel. In Indo-Iranian, laryngeals were also lost if the second vowel was **i* or **u*, producing a monosyllabic diphthong (Lubotsky 1995). In Balto-Slavic, the laryngeals were initially retained before **i* and **u* and eventually produced acute accentuation. The loss of intervocalic laryngeals was therefore an independent innovation in Balto-Slavic and Indo-Iranian.

Grammatical features shared by Indo-Iranian and Balto-Slavic are all archaisms (cf. Kortlandt 2016a). Kortlandt adduces the acc.sg. **h₁mēm* (Skt. *mām*, OCS *mę*) for older **h₁me* (Gr. *ἐμὲ*) ‘me’ as a shared innovation, but this is incorrect. Skt. *mām* is sometimes disyllabic, which is best explained by assuming that it reflects PIE **h₁me* with the Indo-Iranian suffix **-Ham* of Skt. *āham* ‘I’, *t_(u)vām* ‘you’ etc.⁹ OCS *mę*, OPr. *mien* on the other hand, reflect **h₁me* to which the acc.sg. ending **-m* has been added (Olander 2015: 122–3).

The list of shared lexemes provided by Porzig (1954: 164–9) is too short to suggest a closer connection between Indo-Iranian and Balto-Slavic. It includes Skt. *kṛṣṇá-*, Lith. *kišnas*, OCS *črъnъ* ‘black’ < **krs-no-*; Skt. *tucchya-*, Lith. *tūščias*, OCS *toštъ* ‘empty’ < **tusk-io-*; Av. *spənta-*, Lith. *švėntas*, OCS *svętъ* ‘holy’ < **kuen-to-* (possibly with Skt. *śuná-* ‘success’, Hitt. *kunna-* ‘right, favourable’, Duchesne-Guillemin 1947). These are best explained as inherited from PIE. The suffix **-no-* in Skt. *dákṣiṇa-*, Lith. *dėšinas*, OCS *desnъ* ‘right’ < **deks-(i-)no-* may also be an archaism because the suffixes that we find in the other branches, Gr. *δεξιός* < **deks-i-uo-*, Goth. *taihswa* and OIr. *dess* ‘right’ < **deks-uo-*, appear to have been taken over from PIE **lh₂ei-uo-* ‘left’. The lack of medial **-i-* in the Slavic form is not easily explained as an innovation. Lith. *dėšinas* and the Indo-Iranian forms may have been influenced by a lost adverb **deks-i-*, which is often assumed to have existed (Beekes 1994: 90; Stüber 2006).

⁹ I owe this observation to Martin Kümmel.

The discussion above leads to the conclusion that there are hardly any facts that can be better explained if it is assumed that Balto-Slavic was itself part of a larger subgroup of Indo-European.

15.4.2 *Linguistic Contacts of Balto-Slavic and the Depalatalization of Palatovelars*

Although much is known about the linguistic contacts of West Baltic, East Baltic and Slavic when these were already separate branches, language contact dating back to the Balto-Slavic period is more difficult to establish. The part of the Balto-Slavic lexicon that was not derived from inherited Proto-Indo-European material must have been borrowed from unknown contact languages, but these languages are elusive. Many, if not all, non-Indo-European lexemes that can be reconstructed for Proto-Balto-Slavic also have reflexes in other branches of Indo-European, which [Matasović \(2013: 98\)](#) attributes to a lack of direct contact between Balto-Slavic and non-Indo-European languages. The borrowings would have entered Balto-Slavic via an Indo-European intermediate. The main problem of this scenario is that the loanwords in question cannot have been borrowed directly from a known Indo-European language, for phonological reasons. At least one of the contact languages must have been an otherwise lost branch of Indo-European, perhaps the Temematic language argued for by [Holzer \(1989\)](#), cf. the discussion in [Matasović 2013: 77–81](#), [Kortlandt 2016b: 84](#) and [Holzer 2018](#). More than one contact language is perhaps required, for example because the sound changes that would characterize Temematic, if real, are found only in part of the borrowed vocabulary. [Kortlandt \(2018a\)](#) argued for another Indo-European contact language, Venedic, “which contained an older non-Indo-European layer and was part of the Corded Ware horizon.”

There have been attempts to explain certain phonological peculiarities of Balto-Slavic as being due to language contact, but these have not been very successful. This can be illustrated by the so-called *centum* reflexes of the Indo-European palatovelars, the first development on [Matasović’s](#) list cited in [Section 15.2.1](#). See [Hock 2004: 11](#) for a survey of the relevant literature.

The Indo-European palatovelars **k̑*, **g̑* and **g̑ʰ* are in most cases reflected as sibilants in Baltic and Slavic, but both branches also have cases in which the palatovelars became velar occlusives. A detailed study of these cases reveals that the velar reflexes can in no way be regarded as being due to language contact, but must be due to a regular development in certain environments ([Meillet 1894](#); [Kortlandt 2009: 27–32](#); [2013b](#); [Matasović 2005a](#)). This is a priori an attractive scenario, because the words in question look like inherited Baltic and Slavic words in all other respects: there is no other phonetic or morphological reason to think that they might be loanwords and they do not

belong to a part of the lexicon that typically contains loanwords (Čekman 1974: 130–1). Moreover, there is a distribution with regard to the environment in which the velar reflexes are found: they virtually only occur when the following syllable contains a resonant or the semivowel **-u-*. This suggests that the velar reflex was regular before these sounds, in some cases with the additional condition that a back vowel must follow. The original distribution was somewhat obscured by the fact that quite a number of roots regularly obtained variants with sibilant and velar reflexes, depending on the ablaut grade. This variation was generally removed by analogy, unless there was a semantic and/or morphological difference between the variants. Consider the following examples of cognate words, which have both sibilant and velar reflexes:

- OCS *zelenъ*, Ru. *zelënyj* ‘green’, Lith. dial. *želtas* ‘greenish’, Latv. *zēlts* ‘gold’ < **ǵ^hel-*

- Lith. *žālias* ‘green’, OCS *zlato* ‘gold’ < **ǵ^hol-*

- Ru. *žěltij*, Slk. *žltý* ‘yellow’ < **ǵ^hl-*

Lith. *geļtas* ‘yellow’ is a contamination of Proto-Balto-Slavic **dʒelt-*, cf. Latv. *zēlts*, and **gilt-*, cf. Ru. *žěltij*. It is of course arbitrary to assume that Lith. *geļtas* is a borrowing and that all the other forms are inherited.

- Lith. *žardas* ‘rack for drying flax’, Ru. *zoród*, *ozoród* ‘haystack’ < **ǵ^(h)ord-o-*

- OCS *žrědbъ*, Ru. *žerd* ‘pole’ < **ǵ^(h)rd-i-*

- Lith. dial. *šlāvē* ‘honour’, OCS *slovo* ‘word’, *slyšati* ‘to hear’ < **kleu-*, **klūs-*

- Lith. *klausyti* ‘to listen’ < **klous-*

Baltic preserves both *kl-* and *sl-*, while Slavic generalized *sl-*: OCS *slušati* ‘to listen’, *slava* ‘fame’ < **klous-*, **klōu-*. Again, assuming that Lith. *klausyti* is a borrowing is extremely unlikely, if only from a semantic point of view.

- Lith. *šlieti*, dial. *šlinti* ‘to lean’ < **klei-*, **klin-*

- OCS *kloniti se* ‘to bow’, Ru. *klonít* ‘to incline’ < **klon-*.

With **sl-* we find deverbial CS *sloniti se* ‘to lean’, a causative-iterative to **kli-n-*. There is no reason to separate Slavic **klon-* and **slon-* (ÈSSJ 10: 67). PSI. **kloniti* is probably a denominative to **klonъ* ‘inclination’, an *o*-stem derived from **kli-n-* ‘to lean’ (cf. YAv. *-srinaomi*).

- Lith. *šviěsti*, dial. *švitėti*, OCS *světěti se* ‘to shine’ < **k_{uo}iet-*, **k_{ui}t-*

- OCS *cvětъ*, Cz. *květ* ‘flower’ < **k_{uo}oit-*

OCS *světъ*, Ru. *svet* ‘light’ < **k_{uo}it-* is a younger deverbial derivative, while the initial consonant of OCS *cvisti* ‘to bloom’ is analogical after the noun ‘flower’. Latv. *kvitēt* ‘to glimmer’ is identical to Lith. dial. *švitėti*, OCS *světěti se*, but has analogical *k-*.

- Lith. *šėšuras* ‘father-in-law’ < **s_{ue}k_{ur}-*

- OCS *svekry* ‘mother-in-law’ < **s_{ue}k_{ru}-*

PSl. **svekrb* ‘father-in-law’ (not **svekrb*, cf. ORu. *svekrb* and the accent of Ru. *svěkor*, Serb., Cr. *svěkar* instead of †*svekór*, †*svěkar*) is based on **svekry* (Derksen 2008: 475).

- Lith. *akmuõ*, OCS *kamy* ‘stone’ < **h₂ekmō(n)*, -*mon-*

Cf. Skt. *āśman-* ‘stone’; the Slavic forms show metathesis **H ... k > *k ... H* after **k̑ > *k*. Lith. *āšmenys* ‘blade’, which is often considered to be a closely related form with a sibilant reflex of the palatovelar, is much more likely to be an inner-Baltic or Balto-Slavic *men*-stem derived from the root of *aštrūs* ‘sharp’, like many other post-Proto-Indo-European *men*-stems in Baltic (cf. Skardžius 1943: 293–4).

Within a single paradigm, the alternations caused by the depalatalization of palatovelars have not been preserved in the daughter languages; either the velar or the sibilant was generalized (see further Kortlandt 2013b):

- Lith. *širdis*, OCS *srъdьce* ‘heart’ < **krd-*, cf. Lith. *šerdis* ‘core, kernel’, OCS *srěda* ‘middle’ < **kerd-*, OPr. *seyr* < **kēr(d)*
- Lith. *kárvė*, OCS *krava* ‘cow’ < **korh₂-u-*, cf. OPr. *curwis* ‘ox’ < **krh₂-u-*
- Lith. *pėkus*, OPr. *pecku* ‘cattle’, with **-k-* from the oblique cases, cf. Skt. gen. sg. *paśvāḥ* < **pek̑-u-os*.
- OCS *zrěno*, OPr. *syrne* ‘grain’, Lith. *žirnis* ‘pea’ < **grh₂-n-*, cf. OHG *kerno* ‘kernel’ < **gerh₂-n-*.
- Lith. *aštrūs*, OCS *ostrb* ‘sharp’ < **h₂ek̑-ro-*, with **-k̑-* reintroduced from the comparative stem **h₂ek̑-i(e)s-* and/or from derivatives, cf. OCS *osla* ‘whetstone’, *ostbnb* ‘sharp point’, *ostrb* ‘thistle’.

Depalatalization of palatovelars must have occurred in several stages, with e.g. depalatalization before **r* already in Proto-Indo-European (Kortlandt 2013b), but the important point with respect to the Balto-Slavic question is that no uniquely Slavic or Baltic change can be shown to have preceded it and that it is not a contact phenomenon. Explanations of the *centum* reflexes in Balto-Slavic that operate with unverifiable prehistoric dialectal differences or large-scale diffusion from other branches of Indo-European, e.g. in the form of secondary satemization of Balto-Slavic (thus Mottausch 2006) or contact with otherwise unattested Indo-European substrata (thus Andersen 2003: 53–8, 66), simply fail to explain the distribution of the velar reflexes.

We can conclude that our present knowledge of the linguistic contacts of Proto-Balto-Slavic is very limited and confined to evidence from the lexicon.

15.5 The Position of Balto-Slavic

All linguistic evidence points to a Balto-Slavic proto-language that must have existed for a significant period after the disintegration of Proto-Indo-European. All shared innovations could have taken place before the first detectable isoglosses between Baltic and Slavic. Explanations for the data that do not

depart from a single Balto-Slavic proto-language (e.g. [Holzer 2001](#); [Andersen 2003](#)) are unnecessarily complicated and involve additional unfalsifiable dimensions such as shifting prehistoric dialects or otherwise unattested contact languages. The uniformity of this proto-language has often been questioned, as the following two quotations by Petit testify:

Si le balto-slave a existé, ce n'est sûrement pas comme une langue totalement unifiée, mais plutôt comme un groupe de dialectes perméables à la diffusion d'isoglosses.

[If Balto-Slavic has existed, it is surely not as a totally unified language, but rather as a group of dialects susceptible to the diffusion of isoglosses.] ([Petit 2004](#): 35)

No scholar would today seriously reconstruct a proto-language as free of internal variation as Schleicher did for Indo-European, and no scholar, not even the staunchest supporters of a proto-language common to Baltic and Slavic, would dare to write a tale in Balto-Slavic. ([Petit 2018](#): 1971)

I disagree with both statements. Proto-Balto-Slavic – the stage right before the first isoglosses between the three branches arose – may have been dialectally diversified, but this diversity cannot be reconstructed (see [Section 15.3.1](#)). There may have been a “Common Balto-Slavic” period, during which innovations could have affected different subsets of predecessor dialects to West Baltic, East Baltic and Slavic, but the evidence for such a period is limited to the handful of innovations potentially shared by East Baltic and Slavic (see [Section 15.3.2](#)).¹⁰ In fact, the linguistic data do not rule out a scenario in which Proto-Balto-Slavic was a dialect or sociolect that was spoken by a relatively small group of people and that any related dialects or sociolects disappeared without leaving a trace. Because there is at present no compelling positive evidence in favour of internal variation in Proto-Balto-Slavic, we should indeed try to reconstruct a monolithic proto-language that contains the ancestors of all Baltic and Slavic forms and structures that are inherited from Proto-Indo-European as well as the results of the shared innovations of Baltic and Slavic. [Villanueva Svensson \(in press\)](#) rightly remarks that the reconstruction of such a proto-language “can be seen as a powerful heuristic device.” Although it is of course not to be expected that we will ever be able to write a story in Balto-Slavic as well as a speaker of that language would have done, trying to do so would be a very useful way of demonstrating the gaps in our knowledge of Proto-Balto-Slavic (see [Kortlandt 2010](#): 49 for an attempt to render Schleicher’s fable in Proto-Balto-Slavic).

If we take away the innovations that characterize Baltic and Slavic as individual branches, we are left with a language that is both phonologically

¹⁰ Petit’s example of OPr. *irmo* ‘arm’ versus OCS *ramo* ‘shoulder’ can be explained from a Proto-Balto-Slavic ablauting *mn*-stem ([Pronk 2014](#)).

and morphologically still quite close to reconstructed Proto-Indo-European. If the Balto-Slavic proto-language is associated with the (earlier phases of the) Middle Dnieper culture, which seems reasonable, the split between Baltic and Slavic can be dated no later than the beginning of the second millennium BCE. The period of shared innovations would then have been up to 1,500 years, which does not seem to be too short or too long for the number of innovations that must have taken place. After the split, Baltic and Slavic developed independently for over two millennia, which accounts for some of the striking differences between Baltic and Slavic that prompted Meillet to doubt the existence of a shared proto-language in the first place (Rozwadowski 1912: 17–18, 33). This is also the period during which speakers of Baltic and Slavic shifted to a more agriculture-based mode of subsistence, as is shown by their distinct agricultural terminology (Pronk & Pronk-Tiethoff 2018). West and East Baltic remained in each other's vicinity for a longer time, which would explain how they borrowed the same words for certain woodland animals, as mentioned above. Eventually, Baltic and Slavic came into contact again as speakers of Slavic started to move north in the early Middle Ages.

If we go further back in time, we can detect traces of contact between Proto-Balto-Slavic and one or more other languages that appear to be otherwise unknown to us. During the third millennium BCE, Proto-Balto-Slavic would have been spoken by people of the Middle Dnieper culture (see Section 15.1). Balto-Slavic was not part of a larger subgroup of Indo-European. There is insufficient support in the data for a prolonged period in which Proto-Balto-Slavic shared innovations with either Germanic or Indo-Iranian (see Section 15.4.1). This suggests that soon after the dissolution of Proto-Indo-European, the speakers of Proto-Balto-Slavic no longer regularly communicated with the speakers of the ancestors of these other branches, which is best explained by assuming that they had become geographically separated from each other.

References

- Andersen, Henning. 1996. *Reconstructing prehistorical dialects: Initial vowels in Slavic and Baltic*. Berlin: Mouton de Gruyter.
- Andersen, Henning. 2003. Slavic and the Indo-European migrations. In Henning Andersen (ed.), *Language contacts in prehistory: Studies in stratigraphy*, 45–76. Amsterdam: Benjamins.
- Anthony, David W. 2007. *The horse, the wheel, and language*. Princeton, NJ: Princeton University Press.
- Arumaa, Peeter. 1955. Die Verwandtschaftsverhältnisse zwischen Baltisch und Slavisch. *Zeitschrift für Slavische Philologie* 24. 9–28.
- Beekes, Robert S. P. 1994. 'Right', 'left' and 'naked' in Proto-Indo-European. *Orbis* 37. 87–96.

- Beekes, Robert S. P. 2011. *Comparative Indo-European linguistics: An introduction*. 2nd ed., revised and corrected by Michiel de Vaan. Amsterdam: Benjamins.
- Brugmann, Karl. 1897. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. Vol. 1. *Einleitung und Lautlehre*. 2nd ed. Strassburg: Trübner.
- Čekman, V. N. 1974. O refleksax indoevropskix *k̑, *g̑ v balto-slavjanskom jazykovom areale [On the reflexes of Indo-European *k̑, *g̑ in the Balto-Slavic language area]. In T. M. Sudnik (ed.), *Balto-slavjanskije issledovanija [Balto-Slavic studies]*, 116–35. Moscow: Nauka.
- Derksen, Rick. 2002. “Rozwadowski’s change” in Baltic. *Baltu filologija* 11. 5–12.
- Derksen, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.
- Derksen, Rick. 2015. *Etymological dictionary of the Baltic inherited lexicon*. Leiden: Brill.
- Dini, Pietro U. 2014. *Foundations of the Baltic languages*. Vilnius: Vilnius University.
- Duchesne-Guillemin, Jacques. 1947. Études hittites. *Transactions of the Philological Society* 45. 73–91.
- ÈSSJ = O. N. Trubačev & A. F. Žuravlev. 1974–2012. *Ètimologičeskij slovar’ slavjanskix jazykov (A–otbtegnoti (se)) [Etymological dictionary of the Slavic languages (A–otbtegnoti (se))]*. Moscow: Nauka.
- Euler, Wolfram. 2007. Sprachgruppen mit naher Verwandtschaft – Methodenreflexion und -kritik. *Res Balticae* 11. 7–28.
- Fraenkel, Ernst. 1950. *Die baltischen Sprachen: Ihre Beziehungen zu einander und zu den indogermanischen Schwesteridiomen als Einführung in die baltische Sprachwissenschaft*. Heidelberg: Winter.
- Gimbutas, Marija. 1963. *The Balts*. London: Thames and Hudson.
- Girdenis, Aleksas & Albertas Rosinas 1977. Some remarks on the vocalism and morphology of Old Prussian. *General Linguistics* 17. 1–7.
- Golaḅ, Zbigniew. 1992. *The origins of the Slavs: A linguist’s view*. Columbus, OH: Slavica.
- Hill, Eugen. 2016. Phonological evidence for a Proto-Baltic stage in the evolution of East and West Baltic. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 13. 205–32.
- Hirt, Hermann. 1895. Über die mit -m- und -bh- gebildeten Kasussuffixe. *Indogermanische Forschungen* 5. 251–5.
- Hock, Wolfgang. 2004. Baltoslavisch: 1. Teil: Phonologie. *Kratylos* 49. 1–32.
- Hock, Wolfgang. 2005. Baltoslavisch: 2. Teil: Morphonologie, Stammbildung, Flexion. *Kratylos* 50. 1–39.
- Hock, Wolfgang. 2006. Baltoslavisch: 3. Teil: Die baltoslavische Sprachgemeinschaft, Nachträge. *Kratylos* 51. 1–24.
- Holvoet, Axel. 2018. The syntax of Balto-Slavic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*, 2000–11. Berlin: De Gruyter Mouton.
- Holvoet, Axel. 2020. *The middle voice in Baltic*. Amsterdam: Benjamins.
- Holzer, Georg. 1989. *Entlehnungen aus einer bisher unbekannten indogermanischen Sprache im Urslavischen und Urbaltischen*. Vienna: Österreichische Akademie der Wissenschaften.
- Holzer, Georg. 2001. Zur Lautgeschichte des baltisch-slavischen Areals. *Wiener slavistisches Jahrbuch* 47. 33–50.

- Holzer, Georg. 2018. Kritisches und Antikritisches zum “Temematischen”. *Filologija* 71. 61–73.
- Jasanoff, Jay H. 2017. *The prehistory of the Balto-Slavic accent*. Leiden: Brill.
- Kluge, Friedrich. 1926. *Nominale Stammbildung der altgermanischen Dialekte*. 3rd ed. Halle: Niemeyer.
- Kortlandt, Frederik. 2009. *Baltica & Balto-Slavica*. Amsterdam: Rodopi.
- Kortlandt, Frederik. 2010. *Studies in Germanic, Indo-European and Indo-Uralic*. Amsterdam: Rodopi.
- Kortlandt, Frederik. 2011. *Selected writings on Slavic and general linguistics*. Amsterdam: Rodopi.
- Kortlandt, Frederik. 2013a. Balto-Slavic personal pronouns and their accentuation. *Baltistica* 48. 5–11.
- Kortlandt, Frederik. 2013b. Palatovelars before syllabic resonants: another look. *Baltistica* 48. 13–17.
- Kortlandt, Frederik. 2016a. Balto-Slavic and Indo-Iranian. *Baltistica* 51. 355–64.
- Kortlandt, Frederik. 2016b. Baltic, Slavic, Germanic. *Baltistica* 51. 81–6.
- Kortlandt, Frederik. 2016c. Slavic *i*-verbs, imperfect, and *jā*-stem nouns. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 42. 75–81.
- Kortlandt, Frederik. 2018a. The expansion of the Indo-European languages. *Journal of Indo-European Studies* 46. 219–31.
- Kortlandt, Frederik. 2018b. Pedersen’s law and the rise of distinctive tone in Baltic and Slavic. *Baltistica* 53. 293–301.
- Kortlandt, Frederik. 2018c. Proto-Baltic? *Baltistica* 53. 175–85.
- Kortlandt, Frederik. 2018d. Rise and fall of vowel length in Slavic. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 44. 287–97.
- Kroonen, Guus. 2012. Consonant gradation in the Proto-Germanic iteratives. In Benedicte Nielsen Whitehead et al. (eds.), *The sound of Indo-European: Phonetics, phonemics, and morphophonemics*, 263–90. Copenhagen: Museum Tusculanum.
- Kushniarevich, Alena, et al. 2015. Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLoS ONE* 10(9). e0135820.
- Leskien, August. 1876. *Die Declination im Slavisch-Litauischen und Germanischen*. Leipzig: Hirzel.
- Lubotsky, Alexander M. 1995. Reflexes of intervocalic laryngeals in Sanskrit. In Wojciech Smoczyński (ed.), *Kuryłowicz memorial volume*. Vol. 1, 213–33. Cracow: Universitas.
- Matasović, Ranko. 2005a. The centum elements in Balto-Slavic. In Gerhard Meiser & Olav Hackstein (eds.), *Sprachkontakt und Sprachwandel*, 363–74. Wiesbaden: Reichert.
- Matasović, Ranko. 2005b. Toward a relative chronology of the earliest Baltic and Slavic sound changes. *Baltistica* 40. 147–57.
- Matasović, Ranko. 2013. Substratum words in Balto-Slavic. *Filologija* 60. 75–102.
- Mažiulis, Vytautas. 2013. *Prūsų kalbos etimologijos žodynas [Old Prussian etymological dictionary]*. 2nd ed. Vilnius: Mokslo ir enciklopedijų leidybos centras.
- Meillet, Antoine. 1894. De quelques difficultés de la théorie des gutturales indo-européennes. *Mémoires de la Société de Linguistique de Paris* 8. 277–304.

- Meillet, Antoine. 1905. *Études sur l'étymologie et le vocabulaire du vieux slave*. Vol. 2. *Formation des noms*. Paris: Bouillon.
- Meillet, Antoine. 1922. *Les dialectes indo-européens*. 2nd ed. Paris: Champion.
- Meiser, Gerhard. 1993. Zur Funktion des Nasalpräsens im Urindogermanischen. In Gerhard Meiser (ed.), *Indogermanica et Italica: Festschrift für H. Rix*, 280–313. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Melchert, H. Craig. 2012. Luvo-Lycian dorsal stops revisited. In Roman Sukač & Ondřej Šefčík (eds.), *The sound of Indo-European 2: Papers on Indo-European phonetics, phonemics and morphophonemics*, 206–18. Munich: LINCOM.
- Mittnik, Alissa, et al. 2018. The genetic prehistory of the Baltic Sea region. *Nature Communications* 9. 442.
- Mottausch, Karl-Heinz. 2006. Eine neue Lösung für ein altes Problem: Kentum und Satəm. *Historische Sprachforschung* 119. 35–76.
- Nepokupnyj, A. P., et al. 1989. *Obščaja leksika germanskix i balto-slavjanskix jazykov [The shared vocabulary of Germanic and Balto-Slavic]*. Kiev: Naukova dumka.
- Olander, Thomas. 2009. *Balto-Slavic accentual mobility*. Berlin: de Gruyter.
- Olander, Thomas. 2015. *Proto-Slavic inflectional morphology: A comparative handbook*. Leiden: Brill.
- Olander, Thomas. 2019. A complicated relationship: Balto-Slavic accentual mobility as a non-trivial shared innovation. In Iván Igartua et al. (eds.), *Perspectives on language structure and language change*, 359–80. Amsterdam: Benjamins.
- Pedersen, Holger. 1905. Die Nasalpräsentia und der slavische Akzent. *Zeitschrift für vergleichende Sprachforschung* 18. 297–425.
- Pedersen, Holger. 1933. *Études lituaniennes*. Copenhagen: Levin & Munksgaard.
- Petit, Daniel. 2004. Les langues baltiques et la question balto-slave. *Histoire Épistémologie Langage* 26. 7–41.
- Petit, Daniel. 2009. La préhistoire des adjectifs déterminés du baltique et du slave. *Bulletin de la Société de Linguistique de Paris* 104. 311–60.
- Petit, Daniel. 2010. *Untersuchungen zu den baltischen Sprachen*. Leiden: Brill.
- Petit, Daniel. 2018. Balto-Slavic. In Jared S. Klein, Brian D. Joseph & Matthias Fritz (eds.), *Handbook of comparative and historical Indo-European linguistics*. Vol. 3, 1960–73. Berlin: De Gruyter Mouton.
- Pohl, Heinz D. 1992. Die baltoslavische Spracheinheit – areale Aspekte. In Bernd Barschel, Maria Kozianka & Karin Weber (eds.), *Indogermanisch, Slawisch und Baltisch*, 137–64. Munich: Sagner.
- Porzig, Walter. 1954. *Die Gliederung des indogermanischen Sprachgebiets*. Heidelberg: Winter.
- Pronk, Tijmen. 2014. Proto-Indo-European *mn*-stems in Balto-Slavic. In Norbert Oettinger & Thomas Steer (eds.), *Das Nomen im Indogermanischen: Morphologie, Substantiv versus Adjektiv, Kollektivum*, 318–26. Wiesbaden: Reichert.
- Pronk, Tijmen & Saskia Pronk-Tiethoff. 2018. Balto-Slavic agricultural terminology. In Guus Kroonen, James P. Mallory & Bernard Comrie (eds.), *Talking Neolithic*, 278–314. Washington, DC: Institute for the Study of Man.
- Rębała, Krzysztof, et al. 2007. Y-STR variation among Slavs: Evidence for the Slavic homeland in the middle Dnieper basin. *Journal of Human Genetics* 52. 406–14.

- Rieken, Elisabeth. 2010. Das Zeichen <sà> im Hieroglyphen-Luwischen. In Aygül Süel (ed.), *Acts of the VIIth international congress of Hittitology, Çorum, August 25–31, 2008*. Vol. 2, 651–60. Ankara: T. C. Çorum Valiliği.
- Rimantienė, Rimutė. 1992. The Neolithic of the Eastern Baltic. *Journal of World Prehistory* 6, 97–143.
- Rinkevičius, Vytautas. 2015. *Prūsistikos pagrindai [Foundations of Old Prussian studies]*. Vilnius: Vilnius University.
- Rozwadowski, Jan. 1912. O pierwotnym stosunku wzajemnym języków bałtyckich i słowiańskich [On the original relationship between Baltic and Slavic]. *Rocznik Slawistyczny* 5, 1–36.
- Rusanova, I. P., & È. A. Symonovič (eds.). 1993. *Slavjane i ix sosedi v konce I tysjačeletija do n.è. – pervoj polovine I tysjačeletija n.è. [The Slavs and their neighbours at the end of the first millennium BCE to the first half of the first millennium CE]*. Moscow: Nauka
- Schmid, Wolfgang P. 1992. Die Stellung des Baltischen im Kreise der indogermanischen Sprachen. In Bernd Barschel, Maria Kozianka & Karin Weber (eds.), *Indogermanisch, Slawisch und Baltisch*, 201–22. Munich: Sagner.
- Senn, Alfred. 1941. On the degree of kinship between Slavic and Baltic. *The Slavonic Year-Book* 1, 251–65.
- Senn, Alfred. 1970. Slavic and Baltic linguistic relations. In Velta Rūke-Draviņa (ed.), *Donum Balticum: To Professor Christian S. Stang on the occasion of his seventieth birthday 15 March 1970*, 485–93. Stockholm: Almqvist and Wiksell.
- Skardžius, Pranas. 1943. *Lietuvių kalbos žodžių daryba [Lithuanian word formation]*. Vilnius: Lietuvos mokslų akademija.
- Sławski, Franciszek. 1970. Lexikalische Neuerungen im Baltisch-Slawischen. In Velta Rūke-Draviņa (ed.), *Donum Balticum: To Professor Christian S. Stang on the occasion of his seventieth birthday 15 March 1970*, 501–5. Stockholm: Almqvist & Wiksell.
- Stang, Christian Schweigaard. 1966. *Vergleichende Grammatik der baltischen Sprachen*. Oslo: Universitetsforlaget.
- Stang, Christian Schweigaard. 1972. *Lexikalische Sonderübereinstimmungen zwischen dem Slavischen, Baltischen und Germanischen*. Oslo: Universitetsforlaget.
- Stüber, Karin. 2006. Die indogermanischen Ausdrücke für ‘rechts’ und ‘links’. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 3, 61–72.
- Trautmann, Reinhold. 1923. *Baltisch-slavisches Wörterbuch*. Göttingen: Vandenhoeck & Ruprecht.
- Udolph, Jürgen. 1979. *Studien zu slavischen Gewässernamen und Gewässerbezeichnungen: Ein Beitrag zur Frage nach der Urheimat der Slaven*. Heidelberg: Winter.
- Vaillant, André. 1950. *Grammaire comparée des langues slaves*. Vol. 1. *Phonologie*. Paris & Lyon: Klincksieck.
- Vaillant, André. 1958. *Grammaire comparée des langues slaves*. Vol. 2. *Morphologie*. Paris: Klincksieck.
- Villanueva Svensson, Miguel. 2011. Anticausative-inchoative verbs in the northern Indo-European languages. *Historische Sprachforschung* 124, 33–58.
- Villanueva Svensson, Miguel. 2014. On the relationship between West Baltic and East Baltic. In Tatjana Civan, Marija V. Zavjalova & Artūras Jūdžentis (eds.), *Baltai ir*

- slavai: dvasinių kultūrų sankirtos [The Balts and the Slavs: intersections of spiritual cultures]*, 162–76. Vilnius: Versmė.
- Villanueva Svensson, Miguel. 2019. The infinitive in Baltic and Balto-Slavic. *Indo-European Linguistics* 7. 194–221.
- Villanueva Svensson, Miguel. In press. Balto-Slavonic. In Jan Fellerer & Neil Bermel (eds.), *The Oxford guide to the Slavonic languages*. Oxford: Oxford University Press.
- Widmer, Manuel, et al. 2017. NP recursion over time: Evidence from Indo-European. *Language* 93. 799–826.
- Young, Steven. 2017. Balto-Slavic. In Mate Kapović (ed.), *The Indo-European languages*. 2nd ed., 479–85. London: Routledge.

Index

- Albanian; *see also* [Balkan languages](#)
and Armenian; *see* [Armenian and Albanian](#)
and Balto-Slavic, 226–8, 281
and Balto-Slavic (language contact); *see*
 [language contact, Albanian and Balto-Slavic](#)
and Celtic; *see* [Celtic and Albanian](#)
and Germanic; *see* [Germanic and Albanian](#)
and Greek; *see* [Greek and Albanian](#)
and Greek (language contact); *see* [language contact, Greek and Albanian](#)
and Indo-Iranian, 230–1
and Italian (language contact); *see*
 [language contact, Italian and Albanian](#)
and Italic; *see* [Italic and Albanian](#)
and Latin (language contact); *see* [language contact, Albanian and Latin](#)
and Messapic; *see* [Messapic and Albanian](#)
alteuropäisch, 161
analogy; *see* [innovations, morphological](#)
Anatolian
 and Celtic, 149
 and Germanic, 166
 and Indo-Iranian, 263
 and Italic, 131
 and Tocharian, 11, 88
 and western subgroups, 76–7
Anatolian hypothesis, 37, 60, 61
archaisms, shared, 20, 21–6, 66, 102–3, 158, 262–3, 280
Armenian; *see also* [Balkan languages](#)
and Albanian, 12, 216, 228–9
and Germanic; *see* [Germanic and Armenian](#)
and Greek; *see* [Greek and Armenian](#)
and Phrygian; *see* [Phrygian and Armenian](#)
augment, 4, 197, 209, 217, 225

backmutation, 53, 57, 58
Balkan languages, 12, 14, 163, 189, 192, 193, 209, 216–17, 228, 233, 237–8, 239, 241;
 see also [Albanian](#); [Armenian](#); [Greek](#);
 [Macedonian, Ancient](#); [Messapic](#);
 [Phrygian](#)
and Germanic; *see* [Germanic and Balkan languages](#)
and Indo-Iranian, 217
and Tocharian; *see* [Tocharian and Balkan languages](#)
Baltic
 and Slavic, 44, 45, 196, 269–70, 274–9, 286–7
 and Uralic (language contact); *see*
 [language contact, Baltic and Uralic](#)
Balto-Slavic
 and Albanian; *see* [Albanian and Balto-Slavic](#)
 and Germanic; *see* [Germanic and Balto-Slavic](#)
 and Indo-Iranian; *see* [Indo-Iranian and Balto-Slavic](#)
Bartholomae's Law
 Indo-European, 44, 246–7
 Indo-Iranian, 44, 246–7, 257
Bayesian inference, 1, 3, 7, 36, 37, 38, 55, 59, 60–1, 262; *see also* [computational methods](#)
biology, evolutionary, 33–4, 36, 37, 59
bootstrap values, 60–1
Brugmann's Law (Indo-Iranian), 44, 164, 248–9

Celtic
 and Albanian, 229
 and Anatolian; *see* [Anatolian and Celtic](#)
 and Germanic, 11, 12, 149, 161–3
 and Greek, 178
 and Tocharian; *see* [Tocharian and Celtic](#)
centum vs. *satem*, 136, 161, 192, 196, 209, 217, 228–9, 238–40, 257–9, 264, 271, 281–2, 283–5
cognacy databases, 7, 36–8, 59, 272; *see also* [Swadesh lists](#)
common language, definition, 8
compensatory lengthening
 Albanian, 227, 232, 234, 235
 Doric, 188–9

- compensatory lengthening (cont.)
 Eastern Ionic, 185
 Goidelic, 144
 Greek, 187, 232
 Indo-Iranian, 248
 Indo-Tocharian, 137
 Ionic-Attic, 185
- computational methods, 1–3, 4, 9, 10, 14–15, 22, 23–4, 26, 33–48, 52–61, 109, 159, 225, 241, 262; *see also* Bayesian inference
- contact; *see* language contact
- convergence, 26–9, 116–18, 122, 161, 162, 163, 165, 182, 184–5, 189, 190–1, 223, 269
- Čop's Law (Anatolian), 68
- Core (Proto-)Indo-European; *see* Indo-Tocharian
- Cowgill's Law (Greek), 177; *see also* relative chronology
- dating, 2, 37–8, 58–61
 Proto-Anatolian, 75–6
 Proto-Balto-Slavic, 270–1
 Proto-Greek, 189
 Proto-Indo-European, 2, 37–8, 59–61, 90
 Proto-Indo-Tocharian, 90
 Proto-Luwic, 70–1
 Proto-Tocharian, 87–8
- deaspiration
 Balto-Slavic, 272
 Nuristani, 254
 Tocharian, 92
- depalatalisation
 Balto-Slavic, 271, 283–5
 Indo-Iranian, 252
 Nuristani, 254
- dialect continuum, 26–9, 147, 159–60, 184, 275
- dialects in proto-languages, 7, 26–9, 274–6, 286
- DNA; *see* genetics
- Dybo's Shortening
 Celtic, 137
 Germanic, 137
 Italic, 137, 210
- Fortunatov's Law (Indic), 250
- Francis' Law (Greek), 210
- genetics, 38, 61, 122, 270, 271
- Germanic
 and Albanian, 163, 229–30
 and Albanian (language contact); *see* language contact, Germanic and Albanian
 and Anatolian; *see* Anatolian and Germanic and Armenian, 165–6
 and Balkan languages, 163
 and Balto-Slavic, 12–13, 163–5, 279–81
 and Celtic; *see* Celtic and Germanic and Greek, 163, 178
 and Illyrian; *see* Illyrian and Germanic and Italic; *see* Italic and Germanic and Latin (language contact); *see* language contact, Germanic and Latin and Messapic; *see* Messapic and Germanic and Thracian; *see* Thracian and Germanic and Tocharian; *see* Tocharian and Germanic and Venetic; *see* Venetic and Germanic
- glottochronology, 33–4, 35, 262
- Graeco-Armenian; *see* Greek and Armenian
- Graeco-Aryan; *see* Greek and Indo-Iranian
- grammaticalisation, 24, 25, 155
 Baltic, 277
 Balto-Slavic, 273
 Greek, 180
 Greek and Armenian, 212
 Italic, 127
 Luwic, 68
 North Germanic, 157
 Tocharian, 88
- Grassmann's Law
 Armenian, 248
 Greek, 247
 Indo-Iranian, 247–8
 Latin, 116, 248
 Macedonian, Ancient, 191
 Tocharian, 92, 248
- Greek; *see also* Balkan languages
 and Albanian, 12, 193, 225–6, 231–7, 241
 and Albanian (language contact); *see* language contact, Albanian and Greek
 and Ancient Macedonian; *see* Macedonian, Ancient, and Greek
 and Armenian, 12, 14, 45, 193–6, 209–16, 225–6
 and Celtic; *see* Celtic and Greek
 and Germanic; *see* Germanic and Greek
 and Indo-Iranian, 12, 178, 196–7, 260, 262
 and Italic; *see* Italic and Greek
 and Messapic; *see* Messapic and Greek
 and Phrygian; *see* Phrygian and Greek
 and Venetic; *see* Venetic and Greek
- Hirt's Law (Balto-Slavic), 271
- Holtzmann's Law (Germanic), 154
- homeland
 definition, 7–8
 Indo-European, 2, 4, 27, 37, 60, 61; *see also* steppe hypothesis; Anatolian hypothesis

- Illyrian, 161
 - and Albanian, 241
 - and Germanic, 163
- Indic
 - and Iranian, 44, 45, 196, 251–7
 - and Nuristani, 251–7
- Indo-Anatolian; *see* Indo-European
- Indo-Anatolian hypothesis, 1–2, 5, 11, 13, 66, 77–9, 89–90, 168, 261–2
- Indo-Aryan; *see* Indic
- Indo-Hittite hypothesis; *see* Indo-Anatolian hypothesis
- Indo-Iranian
 - and Albanian; *see* Albanian and Indo-Iranian
 - and Anatolian; *see* Anatolian and Indo-Iranian
 - and Balkan languages; *see* Balkan languages and Indo-Iranian
 - and Balto-Slavic, 12–13, 281–2
 - and Greek; *see* Greek and Indo-Iranian
 - and Italo-Celtic; *see* Italo-Celtic and Indo-Iranian
 - and Uralic (language contact); *see* language contact, Indo-Iranian and Uralic
- Indo-Tocharian (Core Indo-European, Nuclear Indo-European, non-Anatolian Indo-European), 66, 77–9, 89–90, 122, 197, 232, 239, 251, 280
- Indo-Tocharian hypothesis, 89–97, 168, 261–2
- innovations, shared, 21–6, 225–6
 - lexical, 22, 23, 54, 91
 - morphological, 23, 25, 43, 45–7, 54, 56
 - phonological, 22, 54
 - syntactic, 23–4
 - weighting, 25
- intermediate proto-language, definition, 8–9
- Iranian
 - and Indic; *see* Indic and Iranian
 - and Nuristani, 251–7
 - and Tocharian (language contact); *see* language contact, Tocharian and Iranian
- Italian
 - and Albanian (language contact); *see* language contact, Albanian and Italian
- Italic
 - and Albanian, 230
 - and Anatolian; *see* Anatolian and Italic
 - and Celtic, 11, 44, 102–8, 131, 149
 - and Germanic, 12, 130, 161–2
 - and Greek, 130–1, 178
 - and Slavic, 131
- Italo-Celtic; *see also* Italic and Celtic
 - and Indo-Iranian, 108–9, 111
- Joseph's Law (Celtic), 139–40
- Joseph's Law, expanded (Brittonic and Gaulish), 147
- Junggrammatiker; *see* neogrammarians
- Kluge's Law (Germanic), 153
- language contact, 20, 24; *see also* substrate
 - Albanian, 223, 225
 - Albanian and Balto-Slavic, 227
 - Albanian and Greek, 225
 - Albanian and Italian, 225
 - Albanian and Latin, 227
 - Baltic and Uralic, 271
 - Balto-Slavic, 283–5
 - Brittonic and Gallo-Romance, 145
 - Germanic, 37
 - Germanic and Albanian, 227
 - Germanic and Latin, 154
 - Indo-Iranian, 12, 258, 264
 - Indo-Iranian and Uralic, 264
 - Italo-Celtic, 110
 - Tocharian, 84
 - Tocharian and Iranian, 84–5, 87, 88
 - Tocharian and Samoyedic, 84
 - Tocharian and Uralic, 84, 87, 91
- laryngeal metathesis
 - Indo-European, 93
- laryngeals, loss
 - Armenian, 204
 - Balto-Slavic, 272, 282
 - Celtic, 137–8
 - Greek, 175, 177
 - Indo-Iranian, 248, 282
 - Italic, 118
- Latin
 - and Albanian (language contact); *see* language contact, Albanian and Latin
 - and Germanic (language contact); *see* language contact, Germanic and Latin
- lengthening; *see* compensatory lengthening
- lenition
 - Anatolian, 66, 67
 - Armenian, 205
 - Brittonic, 146
 - Celtic, 142
 - Greek, 175, 177
 - Old Irish, 144
 - Phrygian and Greek, 192
 - Sabellic, 124
- lexical clock, 59
- lexical data; *see* cognacy databases
- lexicostatistics, 33–4, 35, 36
- Lidén's Law (Balto-Slavic), 271
- Lindeman's Law, 118
- Lusitanian, 22, 111, 135

- Macedonian, Ancient, 12, 193; *see also* **Balkan languages**
 and Greek, 12, 191
 and Phrygian, 191
 and Thracian, 191
 maximum compatibility, 1, 53
 maximum parsimony, 53
 merger, phonological, 22, 54, 56
 Messapic, 10, 14; *see also* **Balkan languages**
 and Albanian, 240–1, 249
 and Germanic, 163
 and Greek, 193
 methodology, 9, 10, 13, 14–15, 18–29, 33–48,
 52–61, 90–1; *see also* **Bayesian inference**;
computational methods; **glottochronology**;
lexicostatistics
 neogrammarians, 19–21, 22, 26–7
 network, 28, 57–8
 non-Anatolian (Proto-)Indo-European; *see*
Indo-Tocharian
 Nuclear (Proto-)Indo-European; *see* **Indo-**
Tocharian
 Nuristani
 and Indic; *see* **Indic and Nuristani**
 and Iranian; *see* **Iranian and Nuristani**
 outgroup analysis, 52–3, 57
 Palaeo-Balkan; *see* **Balkan languages**
 paradigmatisation
 Germanic, 154–5
 parallel development, 45, 53, 54, 57, 58,
 90
 para-proto-language, definition, 8
 perfect phylogeny, 37, 57
 Phrygian, 10, 12, 14, 22, 109, 237; *see also*
Balkan languages
 and Ancient Macedonian; *see* **Macedonian**,
Ancient, and **Phrygian**
 and Armenian, 12, 191, 209, 212, 216, 217
 and Greek, 12, 14, 177, 178, 190, 191–3,
 212, 216, 217
 Pictish, 135
 Pre-Greek; *see* **substrate in Greek**
 Pre-Samnite, 106, 115, 118, 125
 preterite-presents (Germanic), 154, 167–8
 productivity, sustained
 Germanic, 168
 prothetic vowels
 Greek, 188
 Greek and Armenian, 204
 Greek and Phrygian, 192
 Greek, Phrygian and Armenian, 192, 217
 Proto-Indo-European, definition, 8
 proto-language, definition, 7–8
 pruning, 28–9
 quantitative methods; *see* **computational**
methods
 Rask/Grimm's Law (Germanic), 153
 relative chronology, 23, 24–5, 29, 34, 38–41,
 43–7, 58; *see also* **dating**
 Armenian, 203
 Balto-Slavic, 44, 271–2
 Celtic, 136
 Germanic, 54–6, 153
 Greek, 175–6, 182
 Greek and Armenian, 45
 Indo-Iranian, 44
 Italic, 24–5
 Italo-Celtic, 44, 103–4, 108
 Tocharian, 84–5
 retentions; *see* **archaisms**, **shared**
 rooting of tree, 56
 ruki change, 27
 Albanian, 258
 Armenian, 206, 258, 282
 Balto-Slavic, 258, 271
 Indo-Iranian, 249–50, 258
 Luwian, 258–9, 282
 Samoyedic
 and Tocharian (language contact); *see* **lan-**
guage contact, **Tocharian and Samoyedic**
satem vs. *centum*; *see* *centum* vs. *satem*
 Sicel, 115–16, 118
 Slavic
 and Baltic; *see* **Baltic and Slavic**
 and Italic; *see* **Italic and Slavic**
 speech community, 8, 18, 24, 26–9
 Stang's Law (Indo-European), 121
 steppe hypothesis, 37, 60, 61
 subgrouping, history of, 18–21
 substrate; *see also* **language contact**
 in Albanian, 240
 in Armenian, 209
 in Balto-Slavic, 283–5
 in Greek, 180–1
 in Greek and Armenian, 214–15
 in Indo-Iranian, 249
 in northwestern languages, 110
 in Tocharian, 84, 91
 Swadesh lists, 36, 127–8, 272; *see also* **cognacy**
databases
 Tartessian, 135
 thorn clusters (Indo-European), 108, 136, 177,
 232, 239–40, 252

- Thracian
 and Ancient Macedonian; *see* [Macedonian, Ancient, and Thracian](#)
 and Germanic, [163](#)
- Thurneysen's Law (Italic), [104](#)
- Thurneysen–Havet's Law (Italic), [119](#)
- time depth; *see* [dating](#)
- Tocharian
 and Anatolian; *see* [Anatolian and Tocharian](#)
 and Balkan languages, [216](#)
 and Celtic, [149](#)
 and Germanic, [166](#)
 and Uralic (language contact); *see* [language contact, Tocharian and Uralic](#)
- Uralic
 and Baltic (language contact); *see* [language contact, Baltic and Uralic](#)
 and Indo-Iranian (language contact); *see* [language contact, Indo-Iranian and Uralic](#)
 and Tocharian (language contact); *see* [language contact, Tocharian and Uralic](#)
- Venetic, [11](#), [22](#), [106](#), [115](#), [116](#), [117](#), [118](#), [119](#), [120](#), [121](#), [122](#), [123](#), [124](#), [127](#), [128](#), [161](#)
 and Germanic, [130](#), [162](#), [165](#)
 and Greek, [193](#)
- Verner's Law (Germanic), [55](#), [153](#), [156](#)
- Verschärfung*, [158](#), [159–60](#)
- Wackernagel's Law, [149](#)
- wave model, [20](#), [26](#), [28–9](#), [159–60](#), [240–1](#)
- weighting; *see* [innovations, weighting](#)
- Winter's Law
 Albanian, [227](#), [235](#)
 Balto-Slavic, [164](#), [227](#), [271](#), [275–6](#)

